

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2003, 11:24:35 ; Search time 2154 Seconds
(Without alignments)
8079.609 Million cell updates/sec

Title: US-09-920-953-2

Perfect score: 598

Sequence: 1 ggcgtgcctgcagacgcgc.....agccatttcgcaccaagcc 598

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sv.*
- 13: gb.un.*
- 14: gb.vl.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vit.*
- 38: em.sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81.4	13.6	1372	8	CELI410
2	79	13.2	587	3	TETHEMOGP
3	77.6	13.0	1197	8	CELI637P
4	76.6	12.8	481	1	AF475938
5	70.4	11.8	146174	1	D90910
6	53.8	9.0	494	3	TETHEMOGT
7	50.8	8.5	3605	1	MKNATHS
8	50.8	8.5	12792	1	AE010389
9	48.2	8.1	1898	10	MUSDELT
10	48.2	8.1	2330	10	MUSUCRBP
11	48.2	8.1	3041	10	MUSTRANS01
12	48	8.0	234817	10	AL663048
13	47.6	8.0	110000	2	LMFLCHR32_06
14	47.6	8.0	110000	2	LMFLCHR36_07
15	47.4	7.9	137936	2	CNS08CAX
16	47.4	7.9	144778	2	CNS07EPR
17	47.2	7.9	3014	3	AY084205
18	47	7.9	10486	1	BJDNAKJ
19	46.4	7.8	411	1	MSM249386
20	46.4	7.8	1737	1	AF130980
21	46.4	7.8	16563	1	AE007025
22	46.4	7.8	35377	1	MTCY48
23	46.4	7.8	117080	2	AC098840
24	45.8	7.7	135216	8	AC069145
25	45.8	7.7	144724	8	AC078948
26	45.2	7.6	149117	2	AC120527
27	45	7.5	11401	1	D50308
28	44.8	7.5	5520	3	AC084329
29	44.6	7.5	11034	1	AE005055
30	44.4	7.4	154494	2	AP005070
31	44.2	7.4	11509	1	AE012381
32	44	7.4	12172	1	AE012305
33	44	7.4	79390	2	AC020442
34	44	7.4	84653	3	AC004290
35	44	7.4	167350	2	AC129717
36	44	7.4	180814	3	AC099023
37	44	7.4	256764	3	AE003792
38	43.8	7.3	32274	1	SCC8A
39	43.4	7.3	867	1	AF145230
40	43.4	7.3	110000	2	LMFLCHR36_03
41	43.4	7.3	177466	1	AE001826
42	43.2	7.2	13461	14	AF084543
43	43.2	7.2	195859	14	AF281817
44	43	7.2	201271	2	AC124194
45	43	7.2	213050	1	AL646067

ALIGNMENTS

RESULT 1	CELI410	1372 bp	mrna	linear	PLN 16-JUN-1994
LOCUS	C.eugametos				
DEFINITION	X72915				
ACCESSION	X72915.1				
VERSION	GI:437980				
KEYWORDS	haemoglobin; LI410 gene; light-induced expression; nuclear gene.				
SOURCE	Chlamydomonas eugametos.				
ORGANISM	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.				
REFERENCE	1 (bases 1 to 1372)				
AUTHORS	Couture,M., Chamberland,H., St-Pierre,B., Lafontaine,J. and Guertin,M.				
TITLE	Nuclear genes encoding chloroplast hemoglobins in the unicellular				

Chlamydomonadaceae: Chlamydomonas.
 1 (bases 1 to 1197)
 REFERENCE
 AUTHORS Couture,M., Chamberland,H., St-Pierre,B., Lafontaine,J. and Guertin,M.
 TITLE Nuclear genes encoding chloroplast hemoglobins in the unicellular green alga Chlamydomonas eugametos
 JOURNAL Mol. Gen. Genet. 243 (2), 185-197 (1994)
 MEDLINE 94232186
 PUBMED 8177215
 2 (bases 1 to 1197)
 REFERENCE
 AUTHORS Guertin,M.
 TITLE Direct Submission
 JOURNAL Submitted (23-MAR-1993) M. Guertin, Laval University, Department of Biochemistry, Pavillon Vachon Room 3426, Quebec, G1K 7P4, CANADA
 FEATURES
 source Location/Qualifiers
 1..1197
 /organism="Chlamydomonas eugametos"
 /strain="UMTEX9"
 /db_xref="taxon:3053"
 /clone_lib="cDNA library in lambda ZAP"
 92..586
 /gene="LI637"
 92..586
 /gene="LI637"
 /codon_start=1
 /product="LI637"
 /protein_id="CAA51421.1"
 /db_xref="GI:437983"
 /db_xref="SWISS-PROT:Q08753"
 /translation="MMRTVQLRLPCIRAOQOPVRPSTATAAATAPARKCPSS
 LFAKLGREAVEAADVFKIVADPTVSTYFNTDMKVRKQFAPFLAYALGASEW
 LKQDMRTAHKDLVPHLSDFHFQAVARHLSLTTELGVPEPDDITDAMAVYASTRTVELN
 MPQ0"
 BASE COUNT 250 a 353 c 338 g 256 t
 ORIGIN

Query Match 13.0%; Score 77.6; DB 8; Length 1197;
 Best Local Similarity 51.8%; Pred. No. 1.4e-05;
 Matches 204; Conservative 0; Mismatches 184; Indels 6; Gaps 1;
 QY 11 CAGACGCGCCACCAGCAGACAGACGCGGAGCGGATGCGGGCGCAAGAGCTGT 70
 DB 168 CCGCCACTGCGCGCGCTGCCACCGCAGCGCGCTAGGAAGTGCCTGCTCT 227
 QY 71 TTGATGACTGGCGCGCAGAGGATGAAGCTGGCGGTGCACACCTTCTACGATAAG 130
 DB 228 TCGCAAGCTGGCGCGCGGCGGCTTGAAGCAGCGGTGCACAAAGTTTACAAACA 287
 / 131 TCGTGGCTACCGGAGCTGCTGCGCTTCTTCAGTCCCTGCACATGCAAGACAGA 190
 DB 288 TCGTGGCGGACCCACCGCTCAGCAGCTACTTCTCCAAACACAGACATGAAGGTC 347
 QY 191 TGAAGCAGGTCAAGTTTCATGAGCTGCTGTTGGGGAGCAGACCAATACAAGGCG 250
 DB 348 CCAAGCAGTTCGGTCTTCTGCGCTATGCTCTGGTGGTGATCCGAGTGAAGGCA 407
 QY 251 GCATGTACGACGACACGCGCCATCTGTGTAAGGGCCACGCGCTGGACACCGCCACT 310
 DB 408 ACATCGCGACCCGACACAGGACCTGGT-----TCTTCACTGAGTGACGCTCC 461
 QY 311 ACAAGATCAAGCAGTACCTTGGAGAGACCGCTGCAAGAGATGGCGCTCAAGCAGG 370
 DB 462 AGGCGGTAGCGCGCCACTGAGCGACACCTCACTGAGTGGGTGTTCCTCCCTGA 521
 QY 371 TCCAGACCGCGCGGAGTGGGAGTCCACCG 404
 DB 522 TCACTGATCTATGCGCAGTAGTGCGCTCTACCCG 555

RESULT 4
 AF475938
 LOCUS AF475938 481 bp DNA linear BCT 20-FEB-2002

Synechococcus sp. PCC 7002 cyanoglobin (gln) gene, complete cds.
 AF475938
 AF475938.1 GI:18766960
 DEFINITION
 VERSION
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source Location/Qualifiers
 1..481
 /organism="Synechococcus sp. PCC 7002"
 /strain="PCC 7002"
 /db_xref="taxon:32049"
 53..427
 /gene="gln"
 53..427
 /gene="gln"
 /codon_start=1
 /transl_table=11
 /product="cyanoglobin"
 /protein_id="AAL79195.1"
 /db_xref="GI:18766961"
 /translation="MASLYERLGGAAAVDLAVERFYKVLADERNRFFVNTDMAKQK
 QHOKDFMTYAFGGTDRFPGRSMRAHQDLVENAGLTDVHFDIAENLVLTLQELNVSQ
 DLIDEVTVIGSVQHRNDVLR"
 BASE COUNT 129 a 93 c 129 g 130 t
 ORIGIN

Query Match 12.8%; Score 76.6; DB 1; Length 481;
 Best Local Similarity 52.0%; Pred. No. 2.5e-05;
 Matches 172; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
 QY 68 TGTTCATGACCTGGCGCGCAGAGGATGAAGCTGGCGGTTCACACCTTCTACGATA 127
 DB 63 TGTATGAAAGTTAGCGCGCAGCGGCGGATGATCTGGCGGTGAGAAATCTACG 122
 QY 128 AGTGTCTGCTCACCAGGAGCTGCTGCTTCTTCAGTCCCTGAGATGCAAGAGCA 187
 DB 123 AGTTCTGCGCATGAACAGATTAATCGTTTTTGTCTAATGATGATGCGCAACAG 182
 QY 188 AGATGAAGCAGGTCAAGTTTCATGAGCTTCGTGTTTGGCGGAGCAGACCAATACA 247
 DB 183 AGCAACACCAAAAGGATTTTATGACCTATGCTTTTGGGGGCGCAGGATCGATT 242
 QY 248 GAAGCATGTACACGACGACGCGCCATCTGCTCAAGGGCCACGCGCTGGACCC 307
 DB 243 GTTCGATGCGGCTGCCCATCAGGACTTAGTTGAAATGCGGCTCTAACGGATG 302
 QY 308 TTGACAAGATCAAGCAGTACCTTGGAGAGACGCTGCAAGAGATGGCGGTCAAG 367
 DB 303 TCGATGCGATCCCGGAAATCTAGTATTGACCTTCAGGATTAATGTTGTCCCA 362
 QY 368 TGATCCAGCAGCCCGCGGAGTGGTGGAGTC 398
 DB 363 TAATCGATCAAGTCGTCCACCATTTGCGGTT 393

RESULT 5
 D90910
 LOCUS D90910 146174 bp DNA linear BCT 04-JUL-2001
 DEFINITION
 Synechocystis sp. PCC 6803 DNA, complete genome, section:12/27,
 1430419-1576592.
 ACCESSION
 D90910 AB001339 BA000022
 VERSION
 D90910.1 GI:1652956
 KEYWORDS
 Synechocystis sp. PCC 6803 (strain:PCC6803) DNA.
 Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

Qy	428	ACTGCCACCCCAACTGATTTTTCATTAAACCCCAACCCCAACGCTT	468
Db	132420	AGTGAAGTTGAGAATTTTATTTTATGGAGCGCATGTTGCAT	132460
RESULT 6			
TETHMOGT			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
ORGANISM			
SOURCE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
CDS			
BASE COUNT			
ORIGIN			
Query Match 9.0%; Score 53.8; DB 3; Leng			
Best Local Similarity 49.9%; Pred. No. 0.81;			
Matches 173; Conservative 0; Mismatches 162; Ind			
Qy	68	TGTTTGATGACCTGGCGCGCAGAGCGCATGAGCTGGCGGTTGACAC	
Db	36	TTTTTGAAAGCTTGGCGGACAAAGCTGCTATGCATGCACTGTTTCCGCC	
Qy	128	AGGTGCTGCTGACCGCGAGCTGCTGCCCTTCCTCGAGTCCCTCGACAC	
Db	96	AAGCTTCAGCAGATAGTAGAGTCAACATATCTTTAAACACACTAATA	
Qy	188	AGATGAAGCAGGTCAAGTTTCATGAGCTTCGTGTTTGGCGGAGCAGACCC	
Db	156	CTAAGTAATAGGAAGATTTTCTTACTATGCTTTTAGGAGGACCCCAATC	
Qy	248	GAAGCATGTACGACGCACACGCGCCATCTGGTCAAGGGCCACGGGCTGG	
Db	216	AAACATGCTGAAGCTCAT-----AAAGGTATGAAGCTTTT	
Qy	308	TTGACAAGATCAAGCAGTAGTACCTTGGAGAGAGCGCTGCAAGAGATGGCGCG	
Db	264	TTGACGCTATCATTGAGATCTTGGCGGCACACTTAAAGAAATTAGGAG	
Qy	368	TGATCCAGCACGCCCGCGGAGTGGTGGAGTCCACCGCGCAGCAATTT	
Db	324	TTATTGGTGAAGCAGCAGGTCATTGAAACACACTCGTAAAGACTGT	

CDS

1444..2178
 /gene="MK0996"
 /codon_start=1
 /transl_table=11
 /product="Archaea-specific enzyme of the ATP-grasp
 superfamily"
 /protein_id="AA02209.1"
 /db_xref="GI:19887519"
 /translation="MEVEFIRLADEIPECEFLIEGYPGIGMVGVIARHVVTECGAEP
 AFVKIEMPTAAIIVHEGRVYPSAGVFLDDLSFFYAEQADMGRPVGIVNISEVSAELS
 REIGVEICLAGIRAPDLEDESLYYAATSDKAAELFEGIAEPLKGGTITGVSGPLL
 LEGLSDADAVCLLVEPTGTYPPDKAASRVVEALNEALGLSDVSKLEEBAERIAKAV
 EETVRLREOQRMEEBAEPTERFV"
 2240..2629
 /gene="MK0997"
 2240..2629
 /gene="MK0997"
 /codon_start=1
 /transl_table=11
 /product="Uncharacterized conserved protein"
 /protein_id="AA02210.1"
 /db_xref="GI:19887520"
 /translation="MKMLAITGVSRVVEQLLTGGSTQIIIVRSNGNLDVVRLOQGD
 VFVTYVKRELKKGVSQVATVRSVSVVHEHVVTGGDEHFEREFVAARLILNRVSGR
 VREYKDCGVGRREVDEMESAUSEAY"

gene

CDS

2815..3414
 /gene="MK0998"
 2815..3414
 /gene="MK0998"
 /codon_start=1
 /transl_table=11
 /product="Predicted cysteine protease of the
 transglutaminase-like superfamily"
 /protein_id="AA02211.1"
 /db_xref="GI:19887521"
 /translation="MVAYQOGLVAFKLVVYVPLTAVPPVLQIKGGEDELVVRTLA
 LFVSDTHQYDVGAPPVYVPLVRAFAELRCRWGACGTASYMFSYLRLYGIPT
 RLVLIEPTDGTGVSTGTRERERFVLAQVQVYVGHVMTVDPSVGLVGTSDRPVQGH
 RVRPDGLTYTPGVCTYCTGGPLVFWDLTGLVELIV"
 complement(3411..5090)
 /gene="cdc9"
 /note="MK0999"
 complement(3411..5090)
 /gene="cdc9"
 /codon_start=1
 /transl_table=11
 /product="ATP-dependent DNA ligase"
 /protein_id="AA02212.1"
 /db_xref="GI:19887522"
 /translation="MYSSLAEPERLERISSRKAKISLIAQLRQCPEDVVDVVALF
 LANOVFGMDPRDLGIGSKMRKVIATATGSDSEVTELFRLGDLGLTAELLKRRK
 TSLDLSRLPMGEVRETFEKFIAEVEGEGAVKRMRLMGLLARAKPEARYLVROAL
 SELRTGVRESTVEEAIQAAGVSRKLVRAHMLNDLGLVAKVAMTKGEEGLRIDL
 PMRIKPLMAQAARNVKEALAEVGGKGAIEIKLQCARVQVHSDGEVRVYTRIEDVT
 HALPDIIVAEKDCVDADDEFILLEGVAVINPDTGPRPFOELLHRIKRKDIIEVRKEI
 PVELHLFCDLVGDESIVDTPFRERRRLEIIVRECEVMLEVOITDDPKAAEMF
 HRALEMHEGVNADLDANTYPGVGKKMLKVPLETDCVWIGGANGKRRKGLIG
 SYLLAVDENKENLLEVGVTGMDDETLETLTKMFDLIVEESGREVRFKPVVFEV
 EFEDIQSPKYSYSGFALRPLRVDRDLGPDADATIEKVRRIIEEVLQKH"
 5366..6247
 /gene="MK1000"
 5366..6247
 /gene="MK1000"
 /codon_start=1
 /transl_table=11
 /product="Predicted serine/threonine protein kinase"
 /protein_id="AA02213.1"
 /db_xref="GI:19887523"
 /translation="MVVSRVVQALPEMEPDFKVLRALELEMRHRDWPDRLLERT
 GLDEKELGYRLSLRDLMDVYTRRTQTVGYLGYLRPEGYDALRALALVDQVLEGLG
 PEIGVGKEADYLSIGPKAQLAVKFRIRIGTSYTKIKRYREYVYKXKRHISWLVNRL
 TAERFEALLHLYPEGVSVPVPAQNRHLVNRFEGRLEATRVENPEAVLNVRLEE
 YEHALEVGCVVHGDSLQSNIVVEGDDVLLIDHAWHVEVSHPSARELVERDVRNVCDFR

gene

CDS

2815..3414
 /gene="MK0998"
 2815..3414
 /gene="MK0998"
 /codon_start=1
 /transl_table=11
 /product="Predicted cysteine protease of the
 transglutaminase-like superfamily"
 /protein_id="AA02211.1"
 /db_xref="GI:19887521"
 /translation="MVAYQOGLVAFKLVVYVPLTAVPPVLQIKGGEDELVVRTLA
 LFVSDTHQYDVGAPPVYVPLVRAFAELRCRWGACGTASYMFSYLRLYGIPT
 RLVLIEPTDGTGVSTGTRERERFVLAQVQVYVGHVMTVDPSVGLVGTSDRPVQGH
 RVRPDGLTYTPGVCTYCTGGPLVFWDLTGLVELIV"
 complement(3411..5090)
 /gene="cdc9"
 /note="MK0999"
 complement(3411..5090)
 /gene="cdc9"
 /codon_start=1
 /transl_table=11
 /product="ATP-dependent DNA ligase"
 /protein_id="AA02212.1"
 /db_xref="GI:19887522"
 /translation="MYSSLAEPERLERISSRKAKISLIAQLRQCPEDVVDVVALF
 LANOVFGMDPRDLGIGSKMRKVIATATGSDSEVTELFRLGDLGLTAELLKRRK
 TSLDLSRLPMGEVRETFEKFIAEVEGEGAVKRMRLMGLLARAKPEARYLVROAL
 SELRTGVRESTVEEAIQAAGVSRKLVRAHMLNDLGLVAKVAMTKGEEGLRIDL
 PMRIKPLMAQAARNVKEALAEVGGKGAIEIKLQCARVQVHSDGEVRVYTRIEDVT
 HALPDIIVAEKDCVDADDEFILLEGVAVINPDTGPRPFOELLHRIKRKDIIEVRKEI
 PVELHLFCDLVGDESIVDTPFRERRRLEIIVRECEVMLEVOITDDPKAAEMF
 HRALEMHEGVNADLDANTYPGVGKKMLKVPLETDCVWIGGANGKRRKGLIG
 SYLLAVDENKENLLEVGVTGMDDETLETLTKMFDLIVEESGREVRFKPVVFEV
 EFEDIQSPKYSYSGFALRPLRVDRDLGPDADATIEKVRRIIEEVLQKH"
 5366..6247
 /gene="MK1000"
 5366..6247
 /gene="MK1000"
 /codon_start=1
 /transl_table=11
 /product="Predicted serine/threonine protein kinase"
 /protein_id="AA02213.1"
 /db_xref="GI:19887523"
 /translation="MVVSRVVQALPEMEPDFKVLRALELEMRHRDWPDRLLERT
 GLDEKELGYRLSLRDLMDVYTRRTQTVGYLGYLRPEGYDALRALALVDQVLEGLG
 PEIGVGKEADYLSIGPKAQLAVKFRIRIGTSYTKIKRYREYVYKXKRHISWLVNRL
 TAERFEALLHLYPEGVSVPVPAQNRHLVNRFEGRLEATRVENPEAVLNVRLEE
 YEHALEVGCVVHGDSLQSNIVVEGDDVLLIDHAWHVEVSHPSARELVERDVRNVCDFR

gene

CDS

2815..3414
 /gene="MK0998"
 2815..3414
 /gene="MK0998"
 /codon_start=1
 /transl_table=11
 /product="ATP-dependent DNA ligase"
 /protein_id="AA02212.1"
 /db_xref="GI:19887522"
 /translation="MYSSLAEPERLERISSRKAKISLIAQLRQCPEDVVDVVALF
 LANOVFGMDPRDLGIGSKMRKVIATATGSDSEVTELFRLGDLGLTAELLKRRK
 TSLDLSRLPMGEVRETFEKFIAEVEGEGAVKRMRLMGLLARAKPEARYLVROAL
 SELRTGVRESTVEEAIQAAGVSRKLVRAHMLNDLGLVAKVAMTKGEEGLRIDL
 PMRIKPLMAQAARNVKEALAEVGGKGAIEIKLQCARVQVHSDGEVRVYTRIEDVT
 HALPDIIVAEKDCVDADDEFILLEGVAVINPDTGPRPFOELLHRIKRKDIIEVRKEI
 PVELHLFCDLVGDESIVDTPFRERRRLEIIVRECEVMLEVOITDDPKAAEMF
 HRALEMHEGVNADLDANTYPGVGKKMLKVPLETDCVWIGGANGKRRKGLIG
 SYLLAVDENKENLLEVGVTGMDDETLETLTKMFDLIVEESGREVRFKPVVFEV
 EFEDIQSPKYSYSGFALRPLRVDRDLGPDADATIEKVRRIIEEVLQKH"
 5366..6247
 /gene="MK1000"
 5366..6247
 /gene="MK1000"
 /codon_start=1
 /transl_table=11
 /product="Predicted serine/threonine protein kinase"
 /protein_id="AA02213.1"
 /db_xref="GI:19887523"
 /translation="MVVSRVVQALPEMEPDFKVLRALELEMRHRDWPDRLLERT
 GLDEKELGYRLSLRDLMDVYTRRTQTVGYLGYLRPEGYDALRALALVDQVLEGLG
 PEIGVGKEADYLSIGPKAQLAVKFRIRIGTSYTKIKRYREYVYKXKRHISWLVNRL
 TAERFEALLHLYPEGVSVPVPAQNRHLVNRFEGRLEATRVENPEAVLNVRLEE
 YEHALEVGCVVHGDSLQSNIVVEGDDVLLIDHAWHVEVSHPSARELVERDVRNVCDFR

gene

CDS

2815..3414
 /gene="MK0998"
 2815..3414
 /gene="MK0998"
 /codon_start=1
 /transl_table=11
 /product="ATP-dependent DNA ligase"
 /protein_id="AA02212.1"
 /db_xref="GI:19887522"
 /translation="MYSSLAEPERLERISSRKAKISLIAQLRQCPEDVVDVVALF
 LANOVFGMDPRDLGIGSKMRKVIATATGSDSEVTELFRLGDLGLTAELLKRRK
 TSLDLSRLPMGEVRETFEKFIAEVEGEGAVKRMRLMGLLARAKPEARYLVROAL
 SELRTGVRESTVEEAIQAAGVSRKLVRAHMLNDLGLVAKVAMTKGEEGLRIDL
 PMRIKPLMAQAARNVKEALAEVGGKGAIEIKLQCARVQVHSDGEVRVYTRIEDVT
 HALPDIIVAEKDCVDADDEFILLEGVAVINPDTGPRPFOELLHRIKRKDIIEVRKEI
 PVELHLFCDLVGDESIVDTPFRERRRLEIIVRECEVMLEVOITDDPKAAEMF
 HRALEMHEGVNADLDANTYPGVGKKMLKVPLETDCVWIGGANGKRRKGLIG
 SYLLAVDENKENLLEVGVTGMDDETLETLTKMFDLIVEESGREVRFKPVVFEV
 EFEDIQSPKYSYSGFALRPLRVDRDLGPDADATIEKVRRIIEEVLQKH"
 5366..6247
 /gene="MK1000"
 5366..6247
 /gene="MK1000"
 /codon_start=1
 /transl_table=11
 /product="Predicted serine/threonine protein kinase"
 /protein_id="AA02213.1"
 /db_xref="GI:19887523"
 /translation="MVVSRVVQALPEMEPDFKVLRALELEMRHRDWPDRLLERT
 GLDEKELGYRLSLRDLMDVYTRRTQTVGYLGYLRPEGYDALRALALVDQVLEGLG
 PEIGVGKEADYLSIGPKAQLAVKFRIRIGTSYTKIKRYREYVYKXKRHISWLVNRL
 TAERFEALLHLYPEGVSVPVPAQNRHLVNRFEGRLEATRVENPEAVLNVRLEE
 YEHALEVGCVVHGDSLQSNIVVEGDDVLLIDHAWHVEVSHPSARELVERDVRNVCDFR

gene

CDS

RYGVHRHPRREFPENSQ"
 6424..7323
 /gene="MK1001"
 6424..7323
 /gene="MK1001"
 /codon_start=1
 /transl_table=11
 /product="Ferredoxin"
 /protein_id="AA02214.1"
 /db_xref="GI:19887524"
 /translation="MPIEIDMSCLLCEACVAACPTGAIRREDGDMHHCIVCGACVKA
 CPVDALCKLEDEREVDGEKVELKRIAMKPBCKHEDPQCGEVCPOEIMRTEDCYPEL
 RGFVCMCLKCMETCPIDAIGKGVPEKSPPEHPDEDDVYVHPKRCVGTCTCQLQVCP
 TDALEWYPTDAEPKAEENKSWPVYVDPDTMELKQGRKVAKIDPEKCTGCTCLAQVCP
 WGAITAADVPVQSGREKNEIDKCVGCGVCAEVCPEGDLIEVDGVAKVPEKCPACKL
 CERACPVDAISINVSYSERSEME"
 7364..8158
 /gene="mhpb"
 /note="MK1002"
 7364..8158
 /gene="mhpb"
 /codon_start=1
 /transl_table=11
 /product="2-keto-4-pentenolate hydratase hydratase"
 /protein_id="AA02215.1"
 /db_xref="GI:19887525"
 /translation="MRVGTLYVDGTPRPAVFLDDEVHVYDLVPLEFTEVHVSDGLGDL
 DYSVPLSELRIGPPVPPPKIICFGLNYREHVEELREMGMDVSEPMVMKAPTAVI
 GHLDTVKLPREARRVDHELAIVVIGCRKRVSPPEARDVGLGTIIINDVTARDIEKR
 EGOWRAKSYDTFAPLPGWIETETLEPGDGLMELRVNGEVRORATDDMVRDPYELVSE
 TSVMTLEPGDIATGTPPGVGMPEPDKITELETERIGRLVHYVGO"
 complement(8524..9825)
 /gene="MK1003"
 complement(8524..9825)
 /gene="MK1003"
 /codon_start=1
 /transl_table=11
 /product="Predicted DNA-binding protein containing a
 Zn-ribbon"
 /protein_id="AA02216.1"
 /db_xref="GI:19887526"
 /translation="MDWALDDTSDPAGGCTTHAAVLLRAELAEAGAEVPGRPLLVRL
 NPVVPKTRGNAVALPEAPWSVDIEAVLVRLKVVVRKGYPTRGVLVYCEGEP
 VCESVYEAVRRLINPGRVKESVDDDDNVVLEGRGIVGAVAAALGPRVRKDHVEYTFEG
 IYAEKYGTEREVESSIREDFRTPTFDNLDSDGVDLITPMTCPVPLVYGRS
 VEPDLVAPDMIKTRFVVEYEIFESNQATDAHLVRVRLADAEYSNPVLDLTVE
 EPRIPGGHYVVRCEDEEVRVDIAAFPRAPPLTEVVAALHPGDEIRVAGALRPETPK
 HPTVNVKLRVLRLEVRVNPVCGRCRSMKSGRKKGFKCSCGERAPEDSKIGV
 EVPBELVEGTYEAPPVARRHLSKPEYLVELGLLEPSPLSR"
 9852..10046
 /gene="MK1004"
 9852..10046
 /gene="MK1004"
 /codon_start=1
 /transl_table=11
 /product="Uncharacterized protein"
 /protein_id="AA02217.1"
 /db_xref="GI:19887527"
 /translation="MVVGQKILASAVAHLLPFLVGVAGVYIFILTASMKDDEHLRLRGF
 LIIVIGGPTSVLVSIAIVGG"
 10051..11028
 /gene="MK1005"
 10051..11028
 /gene="MK1005"
 /codon_start=1
 /transl_table=11
 /product="Predicted transcriptional regulator containing a
 CTH DNA-binding domain"
 /protein_id="AA02218.1"
 /db_xref="GI:19887528"
 /translation="MRAELCADLEALRAGGHEVYVRIERACFDIFVTRDGRGAYIVKY
 LINADGLRREVAEBELRRISHFLEAVPVVVALKRRHTGPLEKGVVYHRYEVPVLDPLTFA
 RLVEGEPKAVADRGQGVVIRADEVDELDSVSRVRRQLRREGRIYLARAEADVEG

[illegible]

RESULT 13

LMFLCHR32_06/c
WPCOMMENT

Sequence split into 28 fragments LOCUS LMFLCHR32 Accession AL499622

Fragment Name	Begin	End
LMFLCHR32_00	1	110000
LMFLCHR32_01	100001	210000
LMFLCHR32_02	200001	310000
LMFLCHR32_03	300001	410000
LMFLCHR32_04	400001	510000
LMFLCHR32_05	500001	610000
LMFLCHR32_06	600001	710000
LMFLCHR32_07	700001	810000
LMFLCHR32_08	800001	910000
LMFLCHR32_09	900001	1010000
LMFLCHR32_10	1000001	1110000
LMFLCHR32_11	1100001	1210000
LMFLCHR32_12	1200001	1310000
LMFLCHR32_13	1300001	1410000
LMFLCHR32_14	1400001	1510000
LMFLCHR32_15	1500001	1610000
LMFLCHR32_16	1600001	1710000
LMFLCHR32_17	1700001	1810000
LMFLCHR32_18	1800001	1910000
LMFLCHR32_19	1900001	2010000
LMFLCHR32_20	2000001	2110000
LMFLCHR32_21	2100001	2210000
LMFLCHR32_22	2200001	2310000
LMFLCHR32_23	2300001	2410000
LMFLCHR32_24	2400001	2510000
LMFLCHR32_25	2500001	2610000
LMFLCHR32_26	2600001	2710000
LMFLCHR32_27	2700001	277709

```

Query Match      8.0%; Score 47.6; DB 2; Length 110000;
Best Local Similarity 46.2%; Pred. NO. 8.4;
Matches 158; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

```

76 GACCTGGCGGCGCAGAACGCATGAAGTTCGCCGGTTGACACCTTCTACGATAAGTGCTG 135
||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
63280 GACGACGAGGACGCCGCGGCGGCGGACTCTCTCTGCGCGGCGGACGACGACGAGACGCT 63221

QY	136	GCTGACCCGGAGCTGTCGCCCTTCTTCGAGTCCTTGGACATCAAGACGACAAGATGAAG	195
Db	63220	GACGCGGGGAGCTGCTGCTGACGCGCAGACAGCAGGAGCGCGACGCGCGGAG	63161
QY	196	CAGGTCAAATTTCATGAGCTTCGCTTTTGGCGAGCAGACCAATACAGGGCGCAGCATG	255
Db	63160	CTGCTGCTGCTGGACGGCGCAGACAGCAGAGACCGCAGCGCGCGAGCTGCTGCTGCTG	63101
QY	256	TACGACGACAGCGCCCATCTGGTCAAGGGCCACGGCCTGGACCCACCGGCACCTTTGACAA	315
Db	63100	GACGCGCAGACGACGAGGAGCGCCACGCGCGGAGCTGCTGCTGGACGGCGCAGAC	63041
QY	316	ATCAAGCAGTAGCTTTGGAGAGACGCTGCAAGAGATGGCGGTCAACGACGATGTCATCCAG	375
Db	63040	GACGAGGACCGCAGCGCGCGAGCTGCTGCTGGACGGCGCAGACGACGAGGACGCT	62981
QY	376	CACGCCGCGGAGTGGATGCCACCCGCGACGAATTTGAC	417
Db	62980	CACGGCGGGAGCTGCTGCTGGACGGCGCAGCAGCAGGAGC	62939

RESULT 14

LMFLCHR36 07/c

WPCOMMENT

Sequence split into 36 fragments LOCUS LMFLCHR36 Accession AL499624

Fragment Name	Begin	End
LMFLCHR36_00	1	110000
LMFLCHR36_01	100001	210000
LMFLCHR36_02	200001	310000
LMFLCHR36_03	300001	410000
LMFLCHR36_04	400001	510000
LMFLCHR36_05	500001	610000
LMFLCHR36_06	600001	710000
LMFLCHR36_07	700001	810000
LMFLCHR36_08	800001	910000
LMFLCHR36_09	900001	1010000
LMFLCHR36_10	1000001	1110000
LMFLCHR36_11	1100001	1210000
LMFLCHR36_12	1200001	1310000
LMFLCHR36_13	1300001	1410000
LMFLCHR36_14	1400001	1510000
LMFLCHR36_15	1500001	1610000
LMFLCHR36_16	1600001	1710000
LMFLCHR36_17	1700001	1810000
LMFLCHR36_18	1800001	1910000
LMFLCHR36_19	1900001	2010000
LMFLCHR36_20	2000001	2110000
LMFLCHR36_21	2100001	2210000
LMFLCHR36_22	2200001	2310000
LMFLCHR36_23	2300001	2410000
LMFLCHR36_24	2400001	2510000
LMFLCHR36_25	2500001	2610000
LMFLCHR36_26	2600001	2710000
LMFLCHR36_27	2700001	2810000
LMFLCHR36_28	2800001	2910000
LMFLCHR36_29	2900001	3010000
LMFLCHR36_30	3000001	3110000
LMFLCHR36_31	3100001	3210000
LMFLCHR36_32	3200001	3310000
LMFLCHR36_33	3300001	3410000
LMFLCHR36_34	3400001	3510000
LMFLCHR36_35	3500001	3529852

Query Match

Query match 8.0%; Score 47.6; DB 2; Length 110000;
Best Local Similarity 46.2%; Pred. No. 8.4;

Matches 158; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 76 GACCTGGCGCGCCAGAGGCATGAAGCTGGCGGTTGACACCTTCTACGATAAGGTGCTG 135

Db 76978 GACGACGAAGACGGCGGACGGCGGAGCTGCTGCTGGAGCGCGCAGACGACGAGGAGGCC 76919

QY 136 GCTGACCCGAGCTGCTGCCCTCTTTCAGTCCCTGGACATGCAAGACGAGAAGATGAAG 195

ALIGNMENTS

RESULT 3

ABL23790
ID ABL23790 standard; DNA: 9338 BP.
XX
AC ABL23790;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 22843.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1: SEQ ID NO 22843; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 9338 BP; 2756 A; 2046 C; 2018 G; 2518 T; 0 other;
Query Match 7.4%; Score 44; DB 23; Length 9338;
Best Local Similarity 46.0%; Pred. No. 0.22;
Matches 149; Conservative 0; Mismatches 175; Indels 0; Gaps 0;
Oy 95 GCATGAAGCTGGCGTTGACACCTCTACGATAGGTGCTGCTGACCCGGAGCTGTGC 154
Db 7371 GCGTCTCGCTGGCGTTCGAAATCATGCGCAGGATGGTCAGATGTTCCCGTGTGGCCA 7430
Oy 155 CTTCTTCGAGTCCCTGACATGCAAGAGCAGAGATGAAGCTCAAGTTTCATGAGCT 214
Db 7431 ACTACTGTAGTGTGTCTTCGCGAGGAGAGGTGATGGACTGCACGACTACGTTCACT 7490
Oy 215 TCGTGTGTTGGCGGAGCAGACCAATACAAAGGCGGAGAGTGTACGACGACACGCCCATC 274
Db 7491 TGGACGAGGAGTGGCGGCTTAATTCGGTGGACGCGGGTGTGGAAGGAGGACGATCGG 7550
Oy 275 TGGTCAAGGGCCAGCGCTGGACCCACCCGCTTTGACAAGATCAAGCAGTACCTGGAG 334
Db 7551 TCGGCATCGCATGCCATCGCGATCGCGGATCGACGGGGTGTGATGATCTCGGACCCAGGTACC 7610
Oy 335 AGACGCTCGAAGAGATGGCGCTCAAGCAGGATGTGATCCAGCAGCCCGCGAGTGGTGG 394

Db 7611 ATGTCAGCCGAGCGGTGACTGTATGACAGATCAGAGCTATCCGCACACTGCTGAGTTGT 7670
Oy 395 AGTCACCCCGCAGCAATTTGACT 418
Db 7671 AATCTCCGCTTAGCCTATTGTAAT 7694
RESULT 4
AAZ50420
ID AAZ50420 standard; DNA: 3128 BP.
XX
AC AAZ50420;
XX
DT 18-MAY-2000 (first entry)
XX
DE Human wild type EPM2A gene encoding Laforin.
XX
KW EPM2A; laforin; chromosome 6q24; tyrosine phosphatase domain;
KW Lafora's disease; anticonvulsant; mutation; gene therapy; treatment;
KW prevention; human; ds.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT
FT Location/Qualifiers
FT 1..996
FT /*tag= a
FT /product= "Laforin"
FT /note= "Has active catalytic site of a protein tyrosine
FT phosphatase"
FT replace (94, G)
FT /*tag= b
FT mutation
FT replace (146, G)
FT /*tag= c
FT mutation
FT replace (163, T)
FT /*tag= d
FT mutation
FT replace (179, A)
FT /*tag= e
FT mutation
FT replace (234-236, GG)
FT /*tag= f
FT /note= "G at position 235 deleted"
FT replace (322, T)
FT /*tag= g
FT mutation
FT replace (412, T)
FT /*tag= h
FT mutation
FT replace (721, T)
FT /*tag= i
FT mutation
FT replace (799..801, AAAC)
FT /*tag= j
FT /note= "A is inserted at position 800"
FT replace (836, A)
FT /*tag= k
FT mutation
FT replace (878, T)
FT /*tag= l
FT polyA_signal 3026..3031
FT /*tag= m
FT polyA_signal 3095..3100
FT /*tag= n
XX
PN WO200005405-A2.
XX
PD 03-FEB-2000.
XX
XX 20-JUL-1999; 99WO-CA00646.
XX
XX 20-JUL-1998; 98US-0093495.
XX
XX 21-APR-1999; 99US-0130269.
XX
XX (SCHE/) SCHERER S W.
XX (MINA/) MINASSIAN B A.
XX (DELG/) DELGADO-ESCUETA A.
XX (ROUL/) ROULEAU G.
XX

PI Scherer SW, Minassian BA, Delgado-Escueta A, Rouleau G;
 XX WPI; 2000-182718/16.
 DR P-PSDB; AAY44887.
 XX
 PT New polynucleotide encoding tyrosine phosphatase and polypeptide useful
 PT for detecting or treating Lafora's diseases and as a probe or primer to
 PT detect mutations in EPM2A gene
 XX
 XX Claim 3; Fig 13; 59pp; English.
 XX
 CC The present sequence is the human EPM2A gene encoding laforin protein.
 CC It is located on chromosome 6q24 and comprises 4 exons. Laforin
 CC comprises a tyrosine phosphatase domain associated with Lafora's
 CC disease, which is also characterised by deletions or mutations
 CC in EPM2A gene. It has anticonvulsant activity. EPM2A can be used
 CC for detecting the presence of mutations in laforin or EPM2A in samples
 CC and in gene therapy for treating or preventing Lafora's disease.
 XX
 XX Sequence 3128 BP; 876 A; 643 C; 763 G; 846 T; 0 other;
 ~

Query Match 7.1%; Score 42.6; DB 21; Length 3128;
 Best Local Similarity 50.7%; Pred. No. 0.37;
 Matches 102; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 1 GCGCTGCTGACAGCGCGCCACAGACAGACAGACAGCGCGGAGCGGATCGCGGCG 60
 DB 22 GTGGTCCACCCCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81

QY 61 AAGAGCTGTTTATGACCTGGCGCGCGAGAGGATGAAGCTGCGGTTGACACCTTC 120
 DB 82 GAGCTGGGCGGTTGGGAGCGCGCGGTGCGCTCGGCTGAGCGCGCGCGCGCGCG 141

QY 121 TACGATAAGTGTGCTGACCGCGGAGCTGCTGCGCTTCTTCAGTCCCTGGACATGCAA 180
 DB 142 GCGAGCGGGCGCTGCGCTGAGGAGCGCGCGCTGCTGCGGAGGTGGAGCTGGCG 201

QY 181 GAGCAGAAGATGAAGCAGTCTC 201
 DB 202 GCCGAGGAGCGCGCGCAGGAC 222

RESULT 5
 AAH98163
 ID AAH98163 standard; CDNA; 3134 BP.
 XX
 AC AAH98163;
 XX
 12-OCT-2001 (first entry)
 Human EST-derived coding sequence SEQ ID NO: 20.
 Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200154477-A2.
 XX
 PD 02-AUG-2001.
 XX
 XX 25-JAN-2001; 2001WO-US02687.
 XX
 XX 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI

PI Cao Y, Drmanac RA, Zhang J, Werhman T;
 XX WPI; 2001-476164/51.
 DR P-PSDB; AAM23504.
 XX
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 XX Claim 1; Page 195-196; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.
 XX
 XX Sequence 3134 BP; 874 A; 649 C; 759 G; 852 T; 0 other;
 ~

Query Match 7.1%; Score 42.6; DB 22; Length 3134;
 Best Local Similarity 50.7%; Pred. No. 0.37;
 Matches 102; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 1 GCGCTGCTGACAGCGCGCCACAGACAGACAGACAGCGCGGAGCGGATCGCGGCG 60
 DB 22 GTGGTCCACCCCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81

QY 61 AAGAGCTGTTTATGATGACCTGGCGCGCGAGAGGATGAAGCTGCGGTTGACACCTTC 120
 DB 82 GAGCTGGGCGGTTGGGAGCGCGCGGTGCGCTCGGCTGAGCGCGCGCGCGCGCG 141

QY 121 TACGATAAGTGTGCTGACCGCGGAGCTGCTGCGCTTCTTCAGTCCCTGGACATGCAA 180
 DB 142 GCGAGCGGGCGCTGCGCTGAGGAGCGCGCGCTGCTGCGGAGGTGGAGCTGGCG 201

QY 181 GAGCAGAAGATGAAGCAGTCTC 201
 DB 202 GCCGAGGAGCGCGCGCAGGAC 222

RESULT 6
 ABL68348
 ID ABL68348 standard; DNA; 25000 BP.
 XX
 AC ABL68348;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Kidney cancer related gene sequence SEQ ID NO:6685.
 XX
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200194629-A2.
 XX
 PD 13-DEC-2001.
 XX
 XX 30-MAY-2001; 2001WO-US10838.
 XX
 XX 05-JUN-2000; 2000US-209473P.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-233133P.
 PR 18-SEP-2000; 2000US-233617P.
 PR 20-SEP-2000; 2000US-234009P.
 PR 20-SEP-2000; 2000US-234034P.
 PR 20-SEP-2000; 2000US-234052P.
 PR 22-SEP-2000; 2000US-234509P.

PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 28-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236112P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX (AVAL-) AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set
XX
XX Claim 1; SEQ ID 6685; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic
XX activity and can be used in gene therapy. M1 can be used for screening
XX an anti-neoplastic agent, and can be used for producing a product which
XX is the data collected with respect to the anti-neoplastic agent as a
XX result of M1, and the data is sufficient to convey the chemical
XX structure and/or properties of the agent. M1 can be used in the
XX treatment of cancer such as colon, breast, stomach, lung, thyroid,
XX oesophageal, ovarian, kidney, prostate or pancreatic cancer,
XX adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
XX infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
XX carcinoma, papillary carcinoma and Wilm's tumour.
XX
XX Sequence 25000 BP; 5912 A; 6790 C; 6491 G; 5807 T; 0 other;
XX
XX Query Match 7.1%; Score 42.4; DB 24; Length 25000;
XX Best Local Similarity 47.7%; Pred. No. 0.79;
XX Matches 124; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
XX
XX 137 CTGACCCGGAGCTGTGCTTCCTTCGAGTCCCTGCATCGACATGCAAGAGAGATGAAGC 196

Db 22166 CAGACCATGTGCCACCTCTCTCTCCACACAGAGCGCCCATGATGCAGAGGAGCTGAAGA 22225
Qy 197 AGGTCAAGTTTCATGAGCTTCGTGTTTGGCGAGCAGACCAATACAAAGGCCCAAGCATGT 256
Db 22226 AGGAGCAGACACACCGCCACCTGGAGCGCATGAAGAGAACATGGACAGACCATTA 22285
Qy 257 AGGAGCAGACCGCCATCTGTTCAAGGGCCAGCGCTGACACCGCCACTTTGACAGA 316
Db 22286 AGGACCTGCAGCAGCCGCTGGACGAAGCCGAGCATGCCCTCAAGGGCCGCAAGAAGC 22345
Qy 317 TCAAGCAGTACTTGGAGAGAGCGCTGCAAGAGATGGCGTCAAGCAGATGTGATCCAGC 376
Db 22346 AGCTCAGAAGCTGGAAGCGGGTGGGAGCTGGAGATGAGTGGAGCCGAGCAGA 22405
Qy 377 AGCGCCCGGAGTGGTGGAG 396
Db 22406 AGCGCAACGACAGTGGGTG 22425
XX
XX RESULT 7
XX AAX90823
XX ID AAX90823 standard; DNA; 705 BP.
XX AC AAX90823;
XX DT 13-JAN-2000 (first entry)
XX DE DNA encoding human Smad6 protein.
XX KW Smad6 protein; expression; anticoagulant; fibrinolytic function;
XX KW vascular endothelium; modulate; transforming growth factor beta; PAI-1;
XX KW TGF beta; over-expression; plasminogen activator inhibitor 1; stroke;
XX KW thrombomodulin; TM; atherosclerotic lesion; myocardial infarction;
XX KW congestive heart failure; dilated cardiomyopathy; deep venous thrombosis;
XX KW disseminated intravascular thrombosis; sepsis; trauma; septic shock;
XX KW systemic inflammatory response syndrome; multiple organ dysfunction;
XX KW atherosclerotic plaque rupture; anti-metastatic agent; ds.
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX mat_peptide 1..705
XX /tag= a
XX /product= "Smad6 protein"
XX
XX PN W09950296-A1.
XX
XX PD 07-OCT-1999.
XX
XX PF 25-MAR-1999; 99WO-US06595.
XX
XX PR 27-MAR-1998; 98US-0079681.
XX
XX PA (ELIL) LILLY & CO ELI.
XX
XX PI Grinnell BW, Myers LJ, Richardson MA;
XX
XX DR WPI; 1999-610838/52.
XX
XX DR P-FSDB; AAY26343.
XX
XX PT Use of Smad6 and Smad7 for developing products for treating, e.g.
XX PT myocardial infarction and multiple organ dysfunction syndrome -
XX
XX PS Claim 1; Page 49; 60pp; English.
XX
XX CC The present sequence is a DNA encoding human Smad6 protein. Smad6
XX CC mediates the expression of anticoagulant and fibrinolytic functions in
XX CC the vascular endothelium, and also modulates transforming growth factor
XX CC beta (TGF beta) secretion. Over-expression of Smad6 increases the
XX CC secretion of plasminogen activator inhibitor 1 (PAI-1) and suppresses the
XX CC expression of thrombomodulin (TM). Smad6 is predominantly expressed in
XX CC atherosclerotic lesions. It can be used for treating or preventing TGF


```

XX AC AAX15147;
XX DT 26-APR-1999 (first entry)
XX DE
XX DE Nucleotide sequence of the human Smad6 cDNA.
XX KW Smad7; Smad7 minimal promoter; TGF-beta activity;
XX KW cell growth modulation; cancer treatment; fibrosis; Smad6;
XX KW lung cancer; eye defect reduction; mammalian embryo; ss.
XX OS Homo sapiens.
XX PN W09853068-A1.
XX PD 26-NOV-1998.
XX XX 20-MAY-1998; 98WO-US10305.
XX PF 06-MAR-1998; 98US-0077033.
XX PR 20-MAY-1997; 97US-0047221.
XX PR 30-SEP-1997; 97US-0060465.
XX PR 25-FEB-1998; 98US-0075940.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Heldin C, Nakao A, Ten Dijke P;
XX WPI; 1999-070149/06.
XX DT
XX PT New isolated Smad7 nucleic acid molecules - which code for
XX PT polypeptides which inhibit transforming growth factor-beta
XX PT superfamily signalling, used to develop agents for treating, e.g.
XX PT cancer or fibrosis
XX PS
XX PS Example 8; Page 73; 130pp; English.
XX CC The present sequence represents the nucleotide sequence of human Smad6
XX CC cDNA. The specification describes human and murine Smad7 sequences, and
XX CC a Smad7 minimal promoter fragment. The products can be used for
XX CC developing agents for detecting or modulating TGF-beta superfamily
XX CC activities. Such agents can be used for modulating cell growth,
XX CC e.g. to treat cancer or fibrosis. Agents which reduce the expression
XX CC or activity of Smad6 or Smad7 genes can be used for treating lung cancer
XX CC characterised by elevated expression of Smad6 or Smad7 genes. Agents
XX CC which reduce the expression or activity of a Smad7 gene or expression
XX CC product can be used for reducing eye defects in a developing mammalian
XX CC embryo.
XX Y
XX Y Sequence 1491 BP; 213 A; 546 C; 523 G; 209 T; 0 other;

Query Match 7.0%; Score 41.6; DB 20; Length 1491;
Best Local Similarity 52.7%; Pred. No. 0.53;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACGCCCATCTGCTCAAGGCGCCGCTGGACCCGCTTTCACAGATCAAGCAGT 325
DB 1076 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACCTGAGGAGCGCCGAGT 1135
QY 326 ACCTTGGAGACGCTGCAAGAGATGGCGCTCAACGAGATGTGATCCAGCAGCGCGC 385
DB 1136 CGGTGCGGCGAAGCGCGACAGATCGGCTTGGGATCTGCTGACGAGGAGCCGAGG 1195
QY 386 GAGTGGTGGAGTCCACCCGCGACGAATTTGACTTNCCTCAACACTGCCG 434
DB 1196 GCGTGTGGGCTTACAACCGCGGCGAGCACCCCATCTTCTGTAACATCCCC 1244

RESULT 10
AAD36162
ID AAD36162 standard; DNA; 2887 BP.
XX
XX AAD36162;

```

```

XX DT 09-AUG-2002 (first entry)
XX DE Human Smad6 DNA.
XX KW Human; Smad6 protein; antisense; cardiovascular disease; infection;
XX KW inflammation; cancer; therapy; gene; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 937..2427
XX FT /*tag= a
XX FT /product= "Human Smad6 protein"
XX PN W0200228878-A1.
XX PD 11-APR-2002.
XX PF 01-OCT-2001; 2001WO-US30645.
XX PR 04-OCT-2000; 2000US-0679298.
XX PA (ISIS-) ISIS PHARM INC.
XX PI Monia BP, Cowser LM;
XX WPI; 2002-394345/42.
XX P-PSDB; AAE22846.
XX PT Oligonucleotides, useful for the modulation of Smad6 expression in the
XX PT treatment or prophylaxis of e.g. cardiovascular disease, are targeted
XX PT to nucleic acid molecule encoding Smad6
XX PS Example 13; Page 95-98; 110pp; English.
XX CC The invention relates to an antisense oligonucleotide targeted to a
XX CC nucleic acid molecule encoding human Smad6 protein, which specifically
XX CC hybridises with the nucleic acid and inhibits its expression. Antisense
XX CC compounds of the invention are used for inhibiting the expression of
XX CC Smad6 in cells and tissues in the treatment of a disease or condition
XX CC associated with Smad6 such as cardiovascular disease, cancer, infection
XX CC and inflammation. They are also useful in the diagnostics, as research
XX CC reagents, in kits and in antisense therapy. The present sequence is
XX CC human Smad6 DNA.
XX SQ Sequence 2887 BP; 509 A; 899 C; 929 G; 549 T; 1 other;

Query Match 7.0%; Score 41.6; DB 24; Length 2887;
Best Local Similarity 52.7%; Pred. No. 0.65;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACGCCCATCTGCTCAAGGCGCCGCTGGACCCGCTTTCACAGATCAAGCAGT 325
DB 2012 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACCTGAGGAGCGCCGAGT 2071
QY 326 ACCTTGGAGACGCTGCAAGAGATGGCGCTCAACGAGATGTGATCCAGCAGCGCGC 385
DB 2072 CGGTGCGGCGAAGCGCGACAGATCGGCTTGGGATCTGCTGACGAGGAGCCGAG 2131
QY 386 GAGTGGTGGAGTCCACCCGCGACGAATTTGACTTNCCTCAACACTGCCG 434
DB 2132 GCGTGTGGGCTTACAACCGCGGCGAGCACCCCATCTTCTGTAACATCCCC 2180

RESULT 11
AAT36036
ID AAT36036 standard; DNA; 3083 BP.
XX
XX AAT36036;
XX
XX AAT36036;
XX DT 20-NOV-1996 (first entry)
XX

```

DE rchd534 gene differentially expressed in cardiovascular disease.
 XX
 KW Cardiovascular disease; differential expression; target gene;
 KW pathway gene; fingerprint gene; atherosclerosis; ischaemia;
 KW reperfusion; hypertension; restenosis; arterial inflammation;
 KW vector; antibody; diagnosis; gene therapy; drug screening;
 KW rchd534 gene; ss.
 OS
 XX Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1032..1739
 FT /*tag= a
 FT misc_difference 16
 FT /*tag= b
 FT /*note= "the base at position 16 is given as
 FT m in the specification"
 FT misc_difference 30
 FT /*tag= c
 FT /*note= "the base at position 30 is given as
 FT w in the specification"
 FT misc_difference 2911
 FT /*tag= d
 FT /*note= "the base at position 2911 is given as
 FT m in the specification"
 XX
 PN WO9624604-A1.
 XX
 XX 15-AUG-1996.
 XX
 XX 09-FEB-1996; 96WO-US01883.
 XX
 PR 07-JUN-1995; 95US-0485573.
 PR 10-FEB-1995; 95US-0386844.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Falb DA;
 XX
 DR WPI; 1996-384391/38.
 DR P-PSDB; AAW03741.
 XX
 PT New genes differentially expressed in cardiovascular disease - and
 PT related vectors, host cells, proteins and antibodies, for diagnosis,
 PT monitoring, treatment and drug screening
 XX
 PS Claim 1: Fig 35A-B; 200pp; English.
 XX
 CC 4 Novel genes, rchd502 (AAAT36033), rchd523 (AAAT36034), rchd528
 CC (AAAT36035) and rchd534 (AAAT36036), are differentially expressed in
 CC endothelial cells subjected to shear stress. HUVEC cells were exposed
 CC to laminar shear stress, which is thought to be responsible for the
 CC prevalence of atherosclerotic lesions in areas of unusual
 CC circulatory flow. RNAs from treated and control cells were used to
 CC generate cDNA libraries. Differentially expressed bands were
 CC identified by electrophoresis, subcloned and sequenced. The
 CC rchd534 gene product (AAW03741) was also sequenced. The gene was
 CC located on chromosome 15. 4 Other genes (see also AAT36033-36) are
 CC differentially expressed in endothelial cells exposed to interleukin-1.
 CC Detection of these 8 novel genes in excess of normal levels allows the
 CC diagnosis of cardiovascular diseases (CVD). The genes can be used to
 CC generate diagnostic probes, to produce recombinant gene products, to
 CC breed transgenic animal models of CVD, in gene replacement therapy, or
 CC (as antisense, ribozyme or triplex sequences) to treat CVD.
 XX
 SQ Sequence 3083 BP; 805 A; 771 C; 746 G; 758 T; 3 other;
 Query Match 7.0%; Score 41.6; DB 17; Length 3083;
 Best Local Similarity 52.7%; Pred. No. 0.66;
 Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Db 1324 ACCTACCTCAGGCGCGGCTTCTGCTGGCCACGCTCAACCTGCAGCGCAGCGT 1383
 Qy 326 ACCTTGAGAGAGCTCAAGAGATGGCGTCAAGCAGGATGTGATCAGCAGCGCCG 385
 Db 1384 CGGTGGCGCAACGCGGCAAGATCGCTTGGGCATCCCTGCTCAGCAAGAGCGCGACG 1443
 Qy 386 GAGTGTGGAGTCCACCGCGGACGAATTTGACTTNCACCAACCTGGCC 434
 Db 1444 GCGTGTGGCTTACACCGCGGCGGAGCACCCATCTTTCGTAACCTCCC 1492
 RESULT 12
 AAX26275
 ID AAX26275 standard; cDNA; 3083 BP.
 XX
 AC AAX26275;
 XX
 XX 24-MAY-1999 (first entry)
 XX
 DE cDNA sequence of rchd534 gene.
 XX
 KW Fingerprinting gene; rchd502; transmembrane protein; cardiovascular;
 KW fingerprint/target gene; up-regulated; endothelial cell; shear-stress;
 KW atherosclerosis; ischemia; reperfusion; hypertension; restenosis; ds.
 XX
 OS Homo sapiens.
 XX
 XX US5882925-A.
 XX
 PD 16-MAR-1999.
 XX
 PF 09-FEB-1996; 96US-0599654.
 XX
 PR 09-FEB-1996; 96US-0599654.
 PR 10-FEB-1995; 95US-0386844.
 PR 07-JUN-1995; 95US-0485573.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 XX
 XX Falb DA;
 XX
 DR WPI; 1999-214071/18.
 DR P-PSDB; AAW95157.
 XX
 PT New polynucleotides consisting of residues 1-1929 of the rchd502
 PT gene - are differentially expressed in cardiovascular disease
 PT states, and can therefore be used to treat and diagnose
 PT cardiovascular diseases
 XX
 PS Disclosure; Fig 34A-D; 121pp; English.
 XX
 CC The invention relates to a rchd502 target/fingerprint gene encoding a
 CC transmembrane protein. The invention provides cDNAs contained in plasmids
 CC pFCHD502SF (ATCC 69981) and pFCHD502SJ (ATCC 69982) that encode the
 CC rchd502 polypeptide, and are differentially expressed in cardiovascular
 CC disease states. Cultured genetically engineered host cell containing the
 CC rchd502 polynucleotides in operative association with a nucleotide
 CC regulatory element are used for producing a polypeptide rchd502 gene
 CC product. Identifying that the fingerprint/target gene rchd502 is
 CC differentially expressed (up-regulated) by endothelial cells subjected
 CC to shear-stress, provides a tool for the diagnosis and treatment of
 CC cardiovascular disease e.g. atherosclerosis, ischemia/reperfusion,
 CC hypertension, restenosis. The fingerprint gene is useful for testing the
 CC efficacy of candidate drugs in basic research and in clinical trials and
 CC or imaging of a diseased cardiovascular tissue. The gene may also be
 CC used in screening for ligands of target gene product receptor domains, as
 CC well as antagonists of the ligand-receptor interaction.
 XX
 SQ Sequence 3083 BP; 805 A; 770 C; 746 G; 759 T; 3 other;
 Query Match 7.0%; Score 41.6; DB 20; Length 3083;
 Best Local Similarity 52.7%; Pred. No. 0.66;
 Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

OY 266 ACGCCCATCTGCTCAAGGCCACCGCTGGACACCGCCACTTTGACRAGATCAAGCAGT 325
 Db 1324 ACCTACCTCAGGGCAGCGCTTCTGCTGGGCCAGCTCAACCTGGAGCAGCGCAGCGAGT 1383
 OY 326 ACCTTTGGAGAGCGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCGCGCG 385
 Db 1384 CGGTGGCGAAGCGCAGCAGAGATCGGCTTCGGCATCTGCTCAGCAGGAGCGCGAG 1443
 OY 386 GAGTGGTGGAGTCCACCGCGACGAAATTGACTTNCACAACTGCCG 434
 Db 1444 CGGTGGCGCTTACACCGCGGAGCAGCCCATCTTCTGCAACTCCCC 1492

RESULT 13
 AAV81861
 ID AAV81861 standard; cDNA: 3083 BP.
 XX
 AC AAV81861;
 VY
 11-MAR-1999 (first entry)
 DE Human rchd534 encoding cDNA.
 KW Human; cardiovascular disease; atherosclerosis; ischaemia; restenosis;
 KW reperfusion; hypertension; arterial inflammation; diagnosis; rchd528; ds.
 XX Homo sapiens.

Key Location/Qualifiers
 CDS 1032..1739
 FT /*tag= a
 XX US5849578-A.
 PN 15-DEC-1998.
 PD 15-MAR-1996; 96US-0616844.
 PF 09-FEB-1996; 96US-0599654.
 PR 10-FEB-1995; 95US-0386844.
 PR 07-JUN-1995; 95US-0458873.
 PR 15-MAR-1996; 96US-0616844.
 XX (MILL-) MILLENNIUM PHARM INC.
 PA Falb DA;
 PI WPI: 1999-069743/06.
 XX P-PSDB: AAW89301.

PT DNA encoding rchd528 polypeptide - associated with cardiovascular
 PT disease
 XX Example; Fig 34; 122pp; English.

XX The present sequence encodes rchd534 protein. The present invention
 CC describes rchd528 protein. A method has been developed for producing
 CC the rchd528 gene product. The present invention describes methods and
 CC compositions for the treatment and diagnosis of cardiovascular diseases,
 CC including: atherosclerosis; ischaemia; restenosis; reperfusion;
 CC hypertension; and arterial inflammation.
 XX Sequence 3083 BP; 805 A; 770 C; 746 G; 759 T; 3 other;

Query Match
 Best Local Similarity 7.0%; Score 41.6; DB 20; Length 3083;
 Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
 OY 266 ACGCCCATCTGCTCAAGGCCACCGCTGGACACCGCCACTTTGACRAGATCAAGCAGT 325
 Db 1324 ACCTACCTCAGGGCAGCGCTTCTGCTGGGCCAGCTCAACCTGGAGCAGCGCAGCGAGT 1383

OY 326 ACCTTTGAGAGAGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCAGCAGCGCGCG 385
 Db 1384 CGGTGGCGAAGCGCAGCAGAGATCGGCTTCGGCATCTGCTCAGCAGGAGCGCGAG 1443
 OY 386 GAGTGGTGGAGTCCACCGCGACGAAATTGACTTNCACAACTGCCG 434
 Db 1444 CGGTGGCGCTTACACCGCGGAGCAGCCCATCTTCTGCAACTCCCC 1492

RESULT 14
 AAV88579
 ID AAV88579 standard; cDNA: 3083 BP.
 XX
 AC AAV88579;
 XX 05-FEB-2001 (first entry)
 DE Human rchd534 cDNA.
 XX
 KW Human; rchd534 gene; differential expression; HUVEC; shear stress;
 KW endothelial cell; cardiovascular disease; inflammation;
 KW atherosclerosis; antiinflammatory; antiatherosclerotic; diagnosis;
 KW gene therapy; chromosome 15; ss.
 XX Homo sapiens.

Key Location/Qualifiers
 CDS 1032..1739
 FT /*tag= a
 XX US6124433-A.
 PN 26-SEP-2000.
 PD 06-OCT-1997; 97US-0944496.
 PF 09-FEB-1996; 96US-0599654.
 PR 10-FEB-1995; 95US-0386844.
 PR 07-JUN-1995; 95US-0485573.
 XX (BGHM-) BRIGHAM & WOMENS HOSPITAL.
 PA (MILL-) MILLENNIUM PHARM INC.
 XX Gimbrone MA, Falb DA;
 PI WPI: 2000-611017/58.
 DR P-PSDB: AAB19627.

PT Novel isolated rchd502 polypeptides, differentially expressed in
 PT response to endothelial cell shear stress, used for diagnosis,
 PT monitoring clinical trails, and treating cardiovascular diseases such
 PT as ischemia -
 XX Example 9.2; Fig 34A-D; 123pp; English.
 PS The present sequence is that of cDNA corresponding to the coding
 CC region and 5' and 3' regions of the novel human rchd534 gene on
 CC chromosome 15. This gene is differentially expressed (up-regulated)
 CC in response to endothelial cell shear stress, and is not induced by
 CC interleukin-1. The cDNA corresponds to clone pFCHD534, which was
 CC isolated from HUVECs subjected to laminar stress. rchd534 is a
 CC homologue of the Drosophila gene Mothers against decapentaplegic
 CC (MAD). It encodes a 235-amino acid protein (see AAB19627).
 CC up-regulation of the rchd534 gene in a disease state may reflect a
 CC protective role for the gene product in responding to disease.
 CC Alternatively, it may have a causative or exacerbating effect on
 CC the disease state. Modulation of rchd534 gene expression, or the
 CC activity of its gene product, will provide a protective effect.
 CC Knowledge of the gene and its protein product may provide for drugs
 CC with greater specificity for the treatment of inflammation and
 CC atherosclerosis, and be used to monitor clinical trials of drugs in
 CC human patients. rchd534 is 1 of 8 novel human genes of the
 CC invention (see AAV88576-83) characterised as being differentially

CC expressed in cardiovascular disease states, and which are of
CC diagnostic or therapeutic use.

XX Sequence 3083 BP; 805 A; 770 C; 746 G; 759 T; 3 other;

SQ

Query Match 7.0%; Score 41.6; DB 21; Length 3083;

Best Local Similarity 52.7%; Pred. No. 0.66; Mismatches 80; Indels 0; Gaps 0;

Matches 89; Conservative 0;

QY 266 ACCCCCATCTGCTCAAGGGCCAGCGCTGGACCAACCGCCACTTTGACAAGATCAAGCAGT 325

DB 1324 ACCTACTCAGGCGACGGCTTCTGCTGGCGCAGCTCAACCTGGAGCAGCGCAGT 1383

QY 326 ACCTGGAGAGCGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCACGACGCCGCG 385

DB 1384 CGGTGGCGGCAAGCGCAGCAAGATCGGCTTCGGCATCTGCTCAGCAAGGAGCGCGACG 1443

QY 386 GAGTGGTGGAGTCCACCGCGCAGCAATTTGACTTNCACCAACTCGGC 434

DB 1444 GCGTGGGCGCTACAACCGCGGCGACCGCCCATCTTCTGCAACTCCCC 1492

Search completed: April 23, 2003, 12:44:53

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

Job time : 240 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2003, 11:14:45 ; Search time 1605 Seconds
(without alignments)
6034.206 Million cell updates/sec

Title: US-09-920-953-2

Perfect score: 598

Sequence: 1 gcgcgtcgtcagacgcggc.....agccatttcgaccaagcc 598

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	115.2	19.3	470 14	BQ817465 1030064D0
2	115.2	19.3	516 13	BI725456 1031078D1
3	115.2	19.3	521 13	BI530660 1024107G0
4	115.2	19.3	547 13	BI527851 1024085B0
5	115.2	19.3	610 13	BI999469 1031073H0
6	115.2	19.3	667 13	BI720065 1031047D1

7	115.2	19.3	679 12	BG857036
8	115.2	19.3	709 14	BQ817934
9	111.2	18.6	712 12	BG861306
10	110.6	18.5	563 13	BI724643
11	106.4	17.8	373 13	BI532445
12	101.4	17.0	372 14	BQ825667
13	94.2	15.8	350 14	BQ822410
14	67.6	11.3	303 10	AV638207
15	55.8	9.3	444 13	BM446668
16	55.8	9.3	495 13	BM446751
17	55.8	9.3	670 13	BM448292
18	55.8	9.3	692 13	BM448248
19	52.2	8.7	559 13	BM392984
20	52.2	8.7	559 13	BM394452
21	52.2	8.7	594 13	BM393049
22	52.2	8.7	594 13	BM394579
23	52.2	8.7	620 13	BM393013
24	52.2	8.7	620 13	BM394510
25	52.2	8.7	863 13	BM396751
26	51.4	8.6	677 12	BG860929
27	48.2	8.1	631 10	BB635178
28	48.2	8.1	723 13	BI730817
29	47.4	7.9	624 12	BG848400
30	47.4	7.9	631 13	BI717244
31	47.2	7.9	631 10	AB911065
32	46.8	7.8	687 12	BG858042
33	45.8	7.7	573 13	BI726881
34	45.4	7.6	606 13	BJ225569
35	45	7.5	440 12	BF704726
36	44.8	7.5	720 13	BM578510
37	44.4	7.4	647 13	BI305734
38	44.2	7.4	671 13	BI601312
39	43.8	7.3	473 14	BQ167713
40	43.8	7.3	521 14	BQ167660
41	43.8	7.3	557 14	BQ754674
42	43.8	7.3	648 14	BQ469863
43	43.4	7.3	451 13	BI342292
44	43.4	7.3	523 12	BG608682
45	43.4	7.3	550 10	BE030535

ALIGNMENTS

RESULT 1
BQ817465
LOCUS
DEFINITION
1030064D02.y1 C. reinhardtii CC-1690, Deflagellation (normalized),
Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION
BQ817465
VERSION
BQ817465.1 GI:22067311
KEYWORDS
EST.
SOURCE
Chlamydomonas reinhardtii.
ORGANISM
Chlamydomonas reinhardtii.
REFERENCE
1 (bases 1 to 470)
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.

AUTHORS
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J. P., Shrager, J., Silflow, C. and Storn, D.
TITLE
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1030
JOURNAL
Unpublished (2002)
COMMENT
Contact: Charles Hauser
DCMB Box 91000
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
Location/Qualifiers
1..470
/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Deflagellation
(normalized), Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; Deflagellation library, constructed by John Davies
and Jeffrey McDermott, combines cDNAs from CC-1690 cells
which had been re-synthesizing flagella for 15, 30 and 60
min after being deflagellated by pH shock. PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda Zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT 84 a 162 c 142 g 82 t
ORIGIN

Query Match 19.3%; Score 115.2; DB 14; Length 470;
Best Local Similarity 59.5%; Pred. No. 5.2e-18;
Matches 195; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 78 CCTGGCGGCGCAGAGCATGAAGCTGGCGGTTGACACCTCTACGATAGGTGCTGGC 137
DB 111 CATGGGGAGAGCCCGCGGAAAGGCTGTAGAGCTCTCTATGAGCGTATCGGTGC 170

QY 138 TGACCCCGAGCTGCTGCCCTTCTTCGAGTCCCTGGAGCATGAAGAGCAGAGATGAAGCA 197
DB 171 GGATCCGAGCTCGCGCCCTTCTTCGCCAAGCTTGATGATGAGGAGCAGCGCCCAACA 230

QY 198 GGTCAAGTTTCATGAGCTTGTGGTGGCGGAGCAGCAATACAGAGCGCCGAAGCATGTA 257
DB 231 GGTCCCTTTATGACTTACGTGTTCGGCGGCTCGCGGCTTACGAGGCTCGGACCTGGG 290

QY 258 CGAGCGACAGCCCATCTGTCAGGCGCCAGCGGCTTGACACCTGGACCCGACCTTTGACAGAT 317
DB 291 CGCTCCACCGCGCGCTGATCCCGAGCGGCGATGAACCAACCCACCTTTGACCTGGT 350

QY 318 CAAGCAGTACCTTGGAGAGAGCTGCAAGAGATCGGCGCTCAAGCAGGATGTGATCCAGCA 377
DB 351 GGCAGCCACCTGGAGTCCACCTTGCAGAGAGCTGGGGTTCCGCGAGGAGCTCAAGGCCGA 410

QY 378 CGCGCGCGAGTGTGGAGTCCACCCGC 405
DB 411 GGCAATGGCCATCGTGGCATCTGCCCGC 438

SULT 2
LOCUS 725456
DEFINITION 1031078D10.y1 C. reinhardtii CC-1690, Stress II (normalized),
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BI725456
VERSION BI725456.1 GI:15701151
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.

REFERENCE 1 (bases 1 to 516)
AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre
P., McDermott, J. P., Shrager, J., Silflow, C. and Stern, D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1031
JOURNAL Unpublished (2001)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177

Email: chauser@duke.edu.
Location/Qualifiers
Source 1..516
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress II (normalized)
, Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; Stress condition II library, constructed by John
Davies and Jeffrey McDermott, combines cDNAs from CC-1690
cells grown to mid-log phase in TAP (NH4+ - containing)
and shifted to TAP - NO3- (24hrs); H2 production
conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
Sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
Zap II (Stratagene) in the EcoRI (5') and XhoI (3')
sites. pBluescript II SK- plasmids were excised from the
lambda ZAP clones by superinfection with ExAssist
(Stratagene) phage. The library was normalized using
method 4 described in Bonaldo et al., (1996) Genome
Research 6: 791-806."

BASE COUNT 93 a 181 c 147 g 95 t
ORIGIN

Query Match 19.3%; Score 115.2; DB 13; Length 516;
Best Local Similarity 59.5%; Pred. No. 5.4e-18;
Matches 195; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 78 CCTGGCGGCGCAGAGCATGAAGCTGGCGGTTGACACCTCTACGATAGGTGCTGGC 137
DB 149 CATGGGGAGAGCCCGCGGAAAGGCTGTAGAGCTCTCTATGAGCGTATCGGTGC 208

QY 138 TGACCCCGAGCTGCTGCCCTTCTTCGAGTCCCTGGAGCATGAAGAGCAGAGATGAAGCA 197
DB 209 GGATCCCGAGCTCGCGCCCTTCTTCGCCAAGCTTGACATGAAGAGCAGCGCCCAACA 268

QY 198 GGTCAAGTTTCATGAGCTTGTGGTGGCGGAGCAGCAATACAGAGCGCCGAAGCATGTA 257
DB 269 GGTCCCTTTATGACTTACGTGTTCGGCGGCTCGGCGCTTACGAGGCTCGGACCTGGG 328

QY 258 CGAGCGACAGCCCATCTGCTCAAGGCGCCAGCGCTTGACACCTGGACCCGACCTTTGACAAGAT 317
DB 329 CGCTCCACCGCGCGCTGATCCCGAGCAGGCGATGAACCAACCCACCTTTGACCTGGT 388

QY 318 CAAGCAGTACCTTGGAGAGAGCTGCAAGAGATGGGCGTCAAGCAGGATGTGATCCAGCA 377
DB 389 GGCAGCCACCTGGAGTCCACCTTGCAGAGAGCTGGGGTTCCGCGAGGAGCTCAAGGCCGA 448

QY 378 CGCGCGCGAGTGTGGAGTCCACCCGC 405
DB 449 GGCAATGGCCATCGTGGCATCTGCCCGC 476

RESULT 3
LOCUS BI530660
DEFINITION 1024107G03.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BI530660
VERSION BI530660.1 GI:15371234
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.

REFERENCE 1 (bases 1 to 521)
AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre
P., McDermott, J. P., Shrager, J., Silflow, C. and Stern, D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in

KEYWORDS
SOURCE
ORGANISM

EST.
Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
1 (bases 1 to 610)
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre
P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1031
Unpublished (2001)
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

TATURES
source

1..610
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 2lgr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress II (normalized
) Lambda zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; Stress condition II library, constructed by John
Davies and Jeffrey McDermott, combines cDNAs from CC-1690
cells grown to mid-log phase in TAP (NH4+ - containing)
and shifted to TAP - NO3- (24hrs); H2 production
conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
zap II (Stratagene) in the EcoRI (5') and XhoI (3')
sites. pBluescript II SK- plasmids were excised from the
lambda zap clones by superinfection with ExAssist
(Stratagene) phage. The library was normalized using
method 4 described in Bonaldo et al., (1996) Genome
Research 6: 791-806."
Research 6: 791-806." 112 a 202 c 176 g 119 t 1 others

BASE COUNT
ORIGIN

Query Match 19.3%; Score 115.2; DB 13; Length 610;
Best Local Similarity 59.5%; Pred. No. 5.8e-18;
Matches 195; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
/ 78 CCTGGCGCGCAGAGGCGATGAAGCTGGCGGTTGACACCTTCTACGATAGGTCGTGC 137
Db 170 CATGGGGGAGAGCGCGCGTGGAAAGGCTGTAGACCTCTTCTATGAGCGTATCGTTGC 229
Qy 138 TGACCCGGAGCTGCTGCCCTTCTTCGAGTCCCTGGACATGCAAGAGCAGAGATGAAGCA 197
Db 230 GGATCCGAGCTGCGCGCTTCTTCCCAACGTTGCATGAAGAAGCAGCGCCGCAACA 289
Qy 198 GGTCAAGTTATGACTTACGTTGTTGGCGGAGCAGACCAATACAAAGGCGCGAAGCATGTA 257
Db 290 GGTCCCTTTATGACTTACGTTGTTGGCGGCTGGCGGCTACGAGGTCGGACCTGGG 349
Qy 258 CGACGCACAGCCCATCTGGTCAAGGCGCAGCGCTGGACACCCGCCACTTTGACAAGAT 317
Db 350 GCGCTCCACCGCCCTGATCCGCGAGCAGGCGATGAACCAACCACTTTGACCTGGT 409
Qy 318 CAAGCAGTACCTTGGAGAGCGCTCAAGAGATGGCGCTCAAGCAGGATGTGATCCAGCA 377
Db 410 GGCAGCCACCTGGACTCCACCTTCAAGAGCTGGGGGTTGCGCAGAGCTCAAGGCCGA 469
Qy 378 CCGCCCGGAGTGTGGAGTGCACCCGC 405
Db 470 GGCATGGCCATCTGGGATCTGCGCGC 497

RESULT 6
BI720065

LOCUS
DEFINITION
1031047D12.y1 C. reinhardtii CC-1690, Stress II (normalized),
Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
BI720065
ACCESSION
BI720065
VERSION
BI720065.1 GI:15695760
KEYWORDS
EST.
SOURCE
Chlamydomonas reinhardtii.
ORGANISM
Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
1 (bases 1 to 667)
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre
P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1031
Unpublished (2001)
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES
source

1..667
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 2lgr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress II (normalized
) Lambda zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; Stress condition II library, constructed by John
Davies and Jeffrey McDermott, combines cDNAs from CC-1690
cells grown to mid-log phase in TAP (NH4+ - containing)
and shifted to TAP - NO3- (24hrs); H2 production
conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
zap II (Stratagene) in the EcoRI (5') and XhoI (3')
sites. pBluescript II SK- plasmids were excised from the
lambda zap clones by superinfection with ExAssist
(Stratagene) phage. The library was normalized using
method 4 described in Bonaldo et al., (1996) Genome
Research 6: 791-806."
Research 6: 791-806." 122 a 218 c 202 g 124 t 1 others

BASE COUNT
ORIGIN

Query Match 19.3%; Score 115.2; DB 13; Length 667;
Best Local Similarity 59.5%; Pred. No. 6e-18;
Matches 195; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
Qy 78 CTGGCGCGCAGAGGCGATGAAGCTGGCGGTTGACACCTTCTACGATAGGTCGTGC 137
Db 173 CATGGGGGAGAGCGCGCGTGGAAAGGCTGTAGACCTCTTCTATGAGCGTATCGTTGC 232
Qy 138 TGACCCGGAGCTGCTGCCCTTCTTCGAGTCCCTGGACATGCAAGAGCAGAGATGAAGCA 197
Db 233 GGATCCGCGAGCTGCGCGCTTCTTCGCCCAACGTTGACATGAAGAAGCAGCGCCGCAACA 292
Qy 198 GGTCAAGTTATGACTTACGTTGTTGGCGGAGCAGACCAATACAAAGGCGCGAAGCATGTA 257
Db 293 GGTCCCTTTATGACTTACGTTGTTGGCGGCTGGCGGCTACGAGGTCGCGACCTGGG 352
Qy 258 CGACGCACAGCCCATCTGGTCAAGGCGCAGCGCTGGACCAACCCGCCACTTTGACAAGAT 317
Db 353 GCGCTCCACCGCGCGCTGATCCGCGAGCGGCGATGAACCAACCACTTTGACCTGGT 412

Qy	258	CGAGCCACAGCCCCATCTGTGTAAGGGCCACGGCCTGGACACCGCCACCTTTGACAAGAT	317
Db	278	CGCTCCACAGCCGCCCTGTATCCGCGAGCAGGGCATGAACCAACCACCTTTGACCTGGT	337
Qy	318	CAACGACTACTCTGGAGAGCGCTGCAAGAGATGGGCGTCNAGCAGGATGTGATCCAGCA	377
Db	338	GGCAGCCACCTGAGCTCCACCCCTGCAAGAGCTGGGGTTGCGCAGGAGCTCAAGGCCGA	397
Qy	378	CGCCGCGGAGTGGTGAGTCCACCCGC	405
Db	398	GGCAATGCCATCTGGCATCTGCCCGC	425
RESULT 8			
LOCUS	BQ817934	709 bp	mRNA linear EST 01-AUG-2002
DEFINITION	1030067A09.y1 C. reinhardtii CC-1690, Deflagellation (normalized),		
ACCESSION	Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.		
VERSION	BQ817934		
KEYWORDS	BQ817934.1 GI:22067935		
SOURCE	EST.		
ORGANISM	Chlamydomonas reinhardtii.		
REFERENCE	Chlamydomonas reinhardtii.		
AUTHORS	Chlamydomonas reinhardtii		
TITLE	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;		
JOURNAL	Chlamydomonadaceae; Chlamydomonas.		
COMMENT	1 (bases 1 to 709) Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre, P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D. Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1030 Unpublished (2002) Contact: Charles Hauser DCMB Box 91000 Durham, NC 27708-1000 Tel: 919 613 8159 Fax: 919 613 8177 Email: chauser@duke.edu.		
FEATURES			
source	1..709		
	/organism="Chlamydomonas reinhardtii"		
	/strain="CC-1690 wild type mt+ 21gr"		
	/db_xref="taxon:3055"		
	/clone_lib="C. reinhardtii CC-1690, Deflagellation		
	(normalized), Lambda zap II"		
	/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:		
	XhoI; Deflagellation library, constructed by John Davies		
	and Jeffrey McDermott, combines cDNAs from CC-1690 cells		
	which had been re-synthesizing flagella for 15, 30 and 60		
	min after being deflagellated by pH shock. polyA mRNA was		
	purified from each sample, pooled and cDNA synthesized.		
	The cDNA was directionally cloned into lambda zap II		
	(Stratagene) in the EcoRI (5') and XhoI (3') sites.		
	pBluescript II SK- plasmids were excised from the lambda		
	ZAP clones by superinfection with EXASist (Stratagene)		
	phage. The library was normalized using method 4 described		
	in Bonaldo et al., (1996) Genome Research 6: 791-806."		
BASE COUNT	129 a 219 c 231 g		
ORIGIN			
Query Match	19.3%;	Score 115.2;	DB 14; Length 709;
Best Local Similarity	59.3%;	Pred. NO. 6.2e-18;	
Matches 195;	Conservative 0;	Mismatches 133;	Indels 0; Gaps 0;
Qy	78	CTGTGGGGCGCAGAGCGATGAAGCTGGCGGTTGACACCTTCTACGATAAGGTGCTGGC	137
Db	93	CATGGGGGAGAGCCGCCCGTGGAAAGGCTGTAGACGCTCTTCTATGAGCGTATCGTTGC	152
Qy	138	TGACCCGGAGCTGCTGCCCTTCTTCGAGTCCCTCGACATGCAAGACAGCAAGATGAAGCA	197
Db	153	GGATCCCGACGCTCGCGCCCTTCTTCGCGCAACGTTGACATGAAGAAGACGCGCGCAACA	212

BASE COUNT

ORIGIN

Query Match	18.58;	Score	110.6;	DB	13;	Length	563;
Best Local Similarity	57.3%;	Pred. No.	7.6e-17;				
Matches	188;	Conservative	0;	Mismatches	140;	Indels	0;
Gaps	0;						

Qy	78	CCTGGCGGCGCAGAGCATGAAGCTGGCGGTTGACACCTTACGATAGGTCCTGGC	137
Db	170	CATGGGGGAGAGCCCGCTGNAAGGCTGTAGACGCTCTTATGAGNGTATCGTTGC	229
Qy	138	TGACCCGGAGTGTCGTGCCCTTCTTCGAGTCCCTGGACATGCAAGAGCAGAGAATGAAGCA	197
Db	230	GGATCCCGAGCTCGCGNCCTTCTTCGCGCAACGTTGACATGAAGAACAGCGCCGCAACA	289
Qy	198	GGTCAAGTTCTATGAGCTTCGTGTTGGGAGAGCAGACCAATACAAAGGCCCGAAGCATGTA	257
Db	290	GGTGGCCTTTATGNTCTTACGCTGTTTCGCGCGCTTCGGGCGCTTANAAGGTCGCGACCTGGG	349
Qy	258	CGAGCCACAGCCCATCTGGTCAAGGGCCACGGGCTTGACACCACGCCACTTTGACAAGAT	317
	350	CGCCTCCCAACNCGGCTGATCCCGAGCAGGGGATGACNCCACCACCTTTGACCTGGT	409
Qy	318	CAAGCAGTACCTTTGGAGAGACGTGCAAGAGATGGGCTCAAGCAGATGTGATCCAGCA	377
Db	410	GGCAGCCCACTGNACTCCACCTGCAAGAGCTGGGGGTTCCGCGAGGNACTCAAGGCCGA	469
Qy	378	CGCGCCCGGAGTGGTGGATGCCACCCGC	405
Db	470	GGCAATGGCCATCGTGGCATCTGNCCGC	497

RESULT 11	
BI532445	
LOCUS	373 bp mRNA linear EST 29-AUG-2001
DEFINITION	1024121H02.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
ACCESSION	Chlamydomonas reinhardtii cDNA, mRNA sequence.
VERSION	BI532445
KEYWORDS	EST.
SOURCE	BI532445.1 GI:15373019
ORGANISM	Chlamydomonas reinhardtii.
REFERENCE	Chlamydomonas reinhardtii.
AUTHORS	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
TITLE	Chlamydomonadaceae; Chlamydomonas.
JOURNAL	1 (bases 1 to 373)
COMMENT	Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D. Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1024b Unpublished (2001) Contact: Charles Hauser DCMB Box 91000 Duke University Durham, NC 27708-1000 Tel: 919 613 8159 Fax: 919 613 8177 Email: chauser@duke.edu. Location/Qualifiers 1..373
FEATURES	
source	

1. .373
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda zap
II"
/notes="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; This library, constructed by John Davies and Jeffrey
Mcdermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO₂ and HS medium bubbled with 5% CO₂.
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.

BASE COUNT 68 a 129 c 107 q 69 t

Query Match	17.88;	Score 106.4;	DB 13;	Length 373;
Best Local Similarity	61.6%;	Pred. No. 7e-16;		
Matches 170;	Conservative 0;	Mismatches 106;	Indels 0;	Gaps 0;
Qy	78	CTCTGGGGCGCACAAGGCATGAAGCTGGCGGTTTGACACCTTCTACGATATAAGGTCGTGGC	137	
Db	98	CATGGGGGAGAGCCCGCGTGGAAAAGCGTGTAGACGCTCTTCTATGAGCGTATCGGTGC	157	
Qy	138	TGACCCGGAGCTGCTGCCCTTCTTCGAGTCCCTCGGACATGCAAGACAGAGAAGATGAAGCA	197	
Db	158	GGATCCGAGCTCGCGGCCCTCTTCGGCCAAGCTTGACATGAAGAAGACGCGCCGCAACA	217	
Qy	198	GGTCAAGTTTCATGAGCTTCGTGTTTGGCGGAGCAGACCAATACAAAGGCCCGCAAGCATGTA	257	
Db	218	GGTCGCTTTATGACTTACGTGTTTCGGGGCTTCGGGCGCTACGAGGGTCGCGACCTGGG	277	
Qy	258	CGACGCACACCCCATCTGTGTTCAAGGGCCAGGCGCTTGACACACGCCCATTTTGACAAGAT	317	
Db	278	CGCTCCACACCGCGTCTGTATCCGAGCAGGGGCATGAACCACCACCACTTTTGACCTGGT	337	
Qy	318	CAACAGTACCTTGGAGAGCGCTGCAAGAGATGGG	353	
Db	338	GGCAGCCCACTGTGACTCCACCCCTGCAAGAGCTGGG	373	

RESULT 12	BQ825667	372 bp	mrna	linear	EST 01-AUG-2002
LOCUS					
DEFINITION	1030128H02.y1 C. reinhardtii CC-1690, Deflagellation (normalized), Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.				
ACCESSION	BQ825667				
VERSION	BQ825667.1	GI:22076891			
KEYWORDS	EST.				
SOURCE	Chlamydomonas reinhardtii.				
ORGANISM	Chlamydomonas reinhardtii				
REFERENCE	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonas.				
AUTHORS	Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J. P., Shrager, J., Silflow, C. and Stern, D.				
TITLE	Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1030				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: Charles Hauser DCMB Box 91000 Duke University Durham, NC 27708-1000 Tel: 919 613 8159 Fax: 919 613 8177 Email: Chauser@duke.edu.				
FEATURES	Location/Qualifiers				
source	1..372				

1. 372
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21g"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Deflagellation
(normalized), Lambda Zap II"
/note="vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; Deflagellation library, constructed by John Davies
and Jeffrey McDermott, combines cDNAs from CC-1690 cells
which had been re-synthesizing flagella for 15, 30 and 60
min after being deflagellated by pH shock. PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda zap II
(Stratagene) in the EcoRI (5') and XhoRI (3') sites."

pBluescript II SK- plasmids were excised from the lambda
 ZAP clones by superinfection with ExAssist (Stratagene)
 phage. The library was normalized using method 4 described
 in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT	70 a	130 c	103 g	69 t	ORIGIN

Query Match	17.08;	Score 101.4;	DB 14;	Length 372;
Best Local Similarity	60.9%;	Pred. No. 1.2e-14;		
Matches 165;	Conservative	0;	Mismatches 106;	Indels 0; Gaps 0;
QY	78	CTGTGGGGCGGCAGAGCATGAAGCTTGGGGTTTGACACACCTTCTACGATAAGGTGCTGGC	137	
Db	101	CATGGGGGAGAGCCCTCGTGGAAAAGGTGTAGACGCTCTTCTATGAGCGTATCGTTGC	160	
QY	138	TGACCCGAGAGCTGCTGCCCTTCTTCGAGTCCTCGAGCATGCAAGGAGCAGAGATGAGCA	197	
Db	161	GGATCCCGACGTCGCGGCCCTTCTTCGCCAACGTTTGACATGAAGAGGACGCGCCGCAACA	220	
	198	GGTCAAGTTGATGAGCTTCGTGTTTGGCGGAGCAGACCAATACAAAGGCCCAAGCATGTA	257	
	221	GFTCGCCTTTATGACTTTACGTGTTCGGGGGCTTCGGGGCCCTACGAGGGTCGGGACCTGGG	280	
QY	258	CGAGCGCACGCCCATCTGTTCAAGGGCCACGGGCTTGACACCGCCCACTTTTGACAAGAT	317	
Db	281	CGCTCCCAACCGCGCCTGATCCGCGAGCAGGGGCATGAACCAACCACCACTTTTGACCTGGT	340	
QY	318	CAAGCAGTACCTTGGAGAGAGGCTTGCAAGAG	348	
Db	341	GGAGCGCCCACTTGACATCCACCCCTGCAAGAG	371	

RESULT 13
 BQ822410
 LOCUS
 DEFINITION
 1030100E12.y1 C. reinhardtii CC-1690, Deflagellation (normalized),
 Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
 BQ822410 350 bp mRNA linear EST 01-AUG-2002
 BQ822410
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Chlamydomonas reinhardtii.
 Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.
 1 (bases 1 to 350)
 Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
 ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
 Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants. Project: 1030
 Unpublished (2002)
 CONTACT: Charles Hauser
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu.

```

FEATURES             source
Location/Qualifiers
1..350
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Deflagellation
(normalized), Lambda Zap II"
/note="vector: pluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; Deflagellation library, constructed by John Davies
and Jeffrey McDermott, combines cDNAs from CC-1690 cells
which had been re-synthesizing flagella for 15, 30 and 60
min after being deflagellated by pH shock. PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda zap II.
(Stratagene) in the EcoRI (5') and XhoRI (3') sites."

```

pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT	65 a	120 c	100 g	65 t
ORIGIN				

Query Match	15.8%	Score 94.2	DB 14	Length 350
Best Local Similarity	62.6%	Pred. No. 6.8e-13		
Matches 147	Conservative 0	Mismatches 88	Indels 0	Gaps 0
Oy	78	CCTGGGGGCCAGAGGCATGAAGCTGGCGGTTGACACCTTCTACGATAGAAGTGC	137	
Db	112	CATGGGGGAGAGCGCGCGTGGAAAGCGTGTAGACGCTCTTATGAGCGTATCGTTGC	171	
Oy	138	TGACCCGGAGCTGTCGCCCTTCTTCGAGTCCCTGGACATGCAAGAGCAGAAAGATGAAGCA	197	
Db	172	GGATCCGCGAGCTCGCGCCCTTCTTCGCCCAACGTTGACATGAAGAAGCAGCCGCCCAACA	231	
Oy	198	GGTCAAGTTTCATGAGCTTCGTGTTTGGCGGAGCAGACCAATACAAGGCCCAACGATGTA	257	
Db	232	GGTGCCTTTATGACTTACGTGTTCGGGGGCTCGGGGCCCTACGAGGGTTCGGACCTGGG	291	
Oy	258	CGAGCGCACACGCCCATCTGTGTCAGGGCCACGGCCTTGACACCGCCCACTTTTGAC	312	
Db	292	CGCTCCACACCGCGCGCTGATCCGCGAGCAGGGCATGAACCCACCACTTTGAC	346	

RESULT	14
AV638207	
LOCUS	303 bp mRNA linear EST 15-DEC-2000
DEFINITION	AV638207 Chlamydomonas reinhardtii 5% CO ₂ Chlamydomonas reinhardtii cDNA clone HC083c06_r 5', mRNA sequence.
ACCESSION	AV638207
VERSION	AV638207.1 GI:10781527
KEYWORDS	EST.
SOURCE	Chlamydomonas reinhardtii.
ORGANISM	Chlamydomonas reinhardtii Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaeae; Chlamydomonas. 1 (bases 1 to 303) Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y. and Tabata, S. Generation of expressed sequence tags from low-CO ₂ and high-CO ₂ adapted cells of Chlamydomonas reinhardtii DNA Res. 7 (5), 305-307 (2000)
REFERENCE	
AUTHORS	Contact: Erika Asamizu
TITLE	
JOURNAL	
MEDLINE	20539644
COMMENT	

```

FEATURES
  source
    1. 303
      Location/Qualifiers
        .organism="Chlamydomonas reinhardtii"
        /strain="C9"
        /db_xref="taxon:3055"
        /clone="HC083C06_r"
        /clone_lib="Chlamydomonas reinhardtii 5% CO2"
        /note="Vector: phagescriptII SK-; Site_1: EcoRI; Site_2:
        XhoI; The cDNA library was constructed from cells cultured
        in a medium with bubbling air containing 5% carbon
        dioxide"
BASE COUNT      56 a      100 c      84 q      63 t

```

[illegible]

```
Oy 138 TGACCCGAGCTGCTCCCTTCTTCAGTCCCTGGACATGCAAGAGCAGAGATGAAGCA 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 GGATCCGAGCTCGGCCCTTCTTCGCCAACGTTGACATGAAGAGCAGCGCCGCAACA 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 198 GGTCAAGTTCATGAGCTTCGTGTTGGCGGAGCAGACCAATACAAAGGCCGCAAGCATG 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 GCTCGCCCTTATGACTTACGTTCTCGCGGCTCGGCCCTACGAGGTCGCGACCTG 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15
BM446668 444 bp mRNA linear EST 01-APR-2002
LOCUS DSA002F02.55957 An expressed sequence tag database for the
DEFINITION halotolerant green alga, Dunaliella salina Dunaliella salina cDNA
clone DSA002F02.5, mRNA sequence.
ACCESSION BM446668
VERSION BM446668.1 GI:19852240
KEYWORDS EST.
SOURCE Dunaliella salina.
ORGANISM Dunaliella salina.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Dunaliellaceae; Dunaliella.
REFERENCE 1 (bases 1 to 444)
AUTHORS Cushman,J.C.
TITLE An expressed sequence tag database for the halotolerant green alga,
Dunaliella salina
JOURNAL Unpublished (2002)
COMMENT Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T3 20mer
BACKWARD: T7 21mer
Plate: 002 row: F column: 02
Seq primer: T3 20mer
High quality sequence stop: 444.
Location/Qualifiers
1..444
/organism="Dunaliella salina"
/db_xref="taxon:3046"
/clone="DSA002F02"
/clone_lib="An expressed sequence tag database for the
halotolerant green alga, Dunaliella salina"
/tissue_type="Cells, which was adapted in 2.5M NaCl via a
incremental series from 1.7 to 2.0 to 2.25 to 2.5 M NaCl,
were exposed to 3.4 M NaCl for 5 hours"
/cell_type="Green"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Library construction was performed
according to Stratagene's recommended protocol for the
Lambda UniZapXR vector and cDNA synthesis kit."

```
BASE COUNT 102 a 120 c 114 g 108 t
ORIGIN
Query Match 9.3%; Score 55.8; DB 13; Length 444;
Best Local Similarity 52.3%; Pred. No. 0.0022;
Matches 173; Conservative 0; Mismatches 152; Indels 6; Gaps 2;
Oy 67 CTGTTTGATGACCTGGCGGCGCAGAGCATGAAGCTGGCGGTTGACACCTT---CTAC 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88 CTCTACGAACGCTTGGTGGTGCAGCAGCTGTGGAGCGCGCAGTAGACGCTCTTTTACAAC 147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 124 CATAGAGTGTCTGCGGAGCTGCTGGCCTTCTTCGAGTCCCTGGACATGCAAGAG 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148 GACAAGGTTCTTAGGATGAGACTCTCGCACCTTCTTTGGCTCTATTCCCTATGGAGAG 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 184 CAGAAGATGAACAGCTCAAGTTTCATGAGCTTCGTGTTTGGCGGAGCAGACCAATACAAG 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: April 23, 2003, 13:11:27

Job time : 1611 secs


```

; PRIOR APPLICATION NUMBER: US 60/049,990
; PRIOR FILING DATE: 1997-06-13
; PRIOR APPLICATION NUMBER: US 60/053,040
; PRIOR FILING DATE: 1997-07-18
; PRIOR APPLICATION NUMBER: US 60/066,173
; PRIOR FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1280
; TYPE: DNA.
; ORGANISM: Homo sapiens
US-09-096-776B-4

```

Query Match	7.0%	Score 41.6;	DB 4;	Length 1280;
Best Local Similarity	52.7%;	Pred. No. 0.066;		
Matches 89; Conservative	0;	Mismatches 80;	Indels 0;	Gaps
QY 266	ACGCCCATCTGGTCAAGGGCCACGGCTGGACACACGCCCACTTTGACAAGATCAACGAGT	375		
405	ACCTACTCTAAGGCAGGGGTTCTTGCTTGGGCCAGGCTCAACCTTGGAGCAGCGCAGCGAGT	464		
QY 326	ACCTTGGAGAGCGCTGCAAGAGATGGGCTCAAGCAGGATGTGTATCCAGCAGCGCCGCCG	385		
Db 465	CGTGGCGGAACCGCAGCAAGATCGGCTTCGGCATCTGTCTCAGCAGGACCCGACG	524		
QY 386	GAGTGTGGAGTCCACCCGGCAGAAATTGTACTTNGCCAAACACTGGC	434		
Db 525	GCGTGTGGGCTTACAACCGGGGAGCACGCCCATCTTCGTCAACTGCC	573		

RESULT 3
US-09-082-092-9
: Sequence 9, Application US/09082092
: Patent No. 6251628
: GENERAL INFORMATION:
: APPLICANT: Nakao, Atsuhito
: APPLICANT: Moren, Anita
: APPLICANT: Heuchel, Rainer
: APPLICANT: Itoh, Susumu
: APPLICANT: Afrakhte, Mozghan
: APPLICANT: Soucheinytskyi, Serhiy
: APPLICANT: Brodin, Greger
: APPLICANT: Landstrom, Marene
: APPLICANT: Heldin, Nils-Erik
: APPLICANT: Heldip, Carl-Henrik
: APPLICANT: ten Dijke, Peter
: TITLE OF INVENTION: SMAD7 AND USES THEREOF
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
: STREET: 600 Atlantic Avenue
: CITY: Boston
: STATE: MA
: COUNTRY: U.S.A.
: Zip: 02210-2211
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEO for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/082,092
: FILING DATE: 20-MAY-1998
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/047,221
: FILING DATE: 20-MAY-1997
: APPLICATION NUMBER: 60/060,465
: FILING DATE: 30-SEP-1997
: APPLICATION NUMBER: 60/075,940
: FILING DATE: 25-FEB-1998
: APPLICATION NUMBER: 60/077,033

```

: FILING DATE: 06-MAR-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Van Amsterdam, John R.
: REGISTRATION NUMBER: 40,212
: REFERENCE/DOCKET NUMBER: L0461/7032
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-720-3500
: TELEFAX: 617-720-2441
: TELEX:
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1491 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-09-082-092-9

```

[illegible]

```

RESULT 4
US-08-840-767-3
; Sequence 3, Application US/08840767B
; Patent No. 6255464
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Riggins, Gregory J.
; APPLICANT: Thiagalingam, Sam
; TITLE OF INVENTION: MAD-Related Genes In the Human
; FILE REFERENCE: 01107, 05548
; CURRENT APPLICATION NUMBER: US/08/840,767B
; CURRENT FILING DATE: 1997-04-16
; EARLIER APPLICATION NUMBER: 60/015,823
; EARLIER FILING DATE: 1996-04-18
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 3
; LENGTH: 1524
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-840-767-3

```

	Query Match	7.0%	Score 41.6;	DB 4;	Length 1524;
	Best Local Similarity	52.7%;	Pred. No. 0.07;		
	Matches	89; Conservative	0; Mismatches	80; Indels	0; Gaps
QY	266	ACGCCCATCTTGTC	CAAGGGCCACGGCCTGG	ACCACCCGCCACTTTG	CACAAGATCAAGCAGT 325
Db	649	ACCTACCTCAAGG	CAGCGGCTTCGCTTGG	CGCCAGCTCAACCTG	GAGCAGCGCAGCGAGT 708
QY	326	ACCTTGAGAGAGC	CTCAAGAGATGGCGCTCA	AGCAGGATGTGATTC	CAGCACGCCGCGCG 385
Db	709	CGGTGCGCGCA	ACGGCAGCAAGATCG	CGCTTCGSCATCCT	GCTCAGCAAGGAGCCCGAG 768
QY	386	GATGTGTRGAGT	CCACCCGGCAGAAAT	TTGACTTNC	CAACACTGCGC 434
Db	769	GCGTGTGGGCT	CTACAAACCGGGGCG	AGCAGCCCATCTT	CGTCAACTCCGC 817


```

RESULT 8
US-08-599-654-36
; Sequence 36, Application US/08599654
; Patent No. 5882925
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711

```

RESULT 9
US-08-485-573-36
Sequence 36, Application US/08485573
Patent No. 5968770
GENERAL INFORMATION:
APPLICANT: FALB, DEAN A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA

ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,573
FILING DATE: 07-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/386,844
FILING DATE: 10-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16
FEATURE:
NAME/KEY: misc_feature
LOCATION: 30
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2911
US-08-485-573-36

Query Match 7.0%; Score 41.6; DB 2; Length 3083;
Best Local Similarity 52.7%; Pred. No. 0.086;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
OY 266 ACGCCCATCTGTCAGGCGCCAGCGCTGGACACCGCCACCTTTGACAGATCAAGCAGT 325
Db 1324 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACCTGGAGCAGCGCAGT 1383
OY 326 ACCTTGGAGAGCGTGCAGAGATGGCGTCAAGCAGCATGTGATCCAGCAGCGCGCG 385
b 1384 CGTGGCGGCAACGCCACCAAGATCGGCTTCGGCATCTGCTCAGCAAGAGCGCGAGC 1443
OY 386 GAGTGTGAGTCCACCGCGGACGAATTTGACTTNCACCAACTGCGC 434
Db 1444 GCGTGTGGCCTACAAACCGCGGAGCACCCTCATCTTCGTAACCTCCCC 1492

RESULT 10
US-08-944-868A-36
Sequence 36, Application US/08944868A
Patent No. 6018025
GENERAL INFORMATION:
APPLICANT: FALB, DEAN A
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,868A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/599,654
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/386,844
FILING DATE: 10-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: CORUZZI, LAURA A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-041
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16
FEATURE:
NAME/KEY: misc_feature
LOCATION: 30
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2911
US-08-944-868A-36

Query Match 7.0%; Score 41.6; DB 3; Length 3083;
Best Local Similarity 52.7%; Pred. No. 0.086;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
OY 266 ACGCCCATCTGTCAGGCGCCAGCGCTGGACACCGCCACCTTTGACAGATCAAGCAGT 325
Db 1324 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACCTGGAGCAGCGCAGT 1383
OY 326 ACCTTGGAGAGCGTGCAGAGATGGCGTCAAGCAGCATGTGATCCAGCAGCGCGCG 385
Db 1384 CGTGGCGGCAACGCCACCAAGATCGGCTTCGGCATCTGCTCAGCAAGAGCGCGAGC 1443
OY 386 GAGTGTGAGTCCACCGCGGACGAATTTGACTTNCACCAACTGCGC 434
Db 1444 GCGTGTGGCCTACAAACCGCGGAGCACCCTCATCTTCGTAACCTCCCC 1492

RESULT 11
US-08-944-423A-36
Sequence 36, Application US/08944423A
Patent No. 6020463
GENERAL INFORMATION:
APPLICANT: FALB, DEAN A
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS 95
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,423A
FILING DATE: 06-OCT-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,654
FILING DATE: 09-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/485,573
FILING DATE: JUN-07-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/386,844
FILING DATE: 10-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: CORUZZI, LAURA A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-105
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16
FEATURE:
NAME/KEY: misc_feature
LOCATION: 30
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2911
US-08-944-423A-36

```

[illegible]

RESULT 12
US-08-925-743-36
; Sequence 36, Application US/08925743
; Patent No. 6054558
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas

CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/925,743
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/485,573
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16
FEATURE:
NAME/KEY: misc_feature
LOCATION: 30
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2911
US-08-925-743-36

	Query Match	7.0%	Score 41.6;	DB 3;	Length 3083;
	Best Local Similarity	52.7%;	Pred. NO. 0.086;		
	Matches 89;	Conservative 0;	Mismatches 80;	Indels 0	
Qy	266	ACGCCCATCTGTC	AAGGCCACGGCCT	GAGCACACC GCCACTTTG	CACAAGATCAAG
Db	1324	ACCTTACCTCAGGC	ACGGGCTTGCTGCT	TGGGCAGCTCAA	CCCTGGAGCGCCACGC
Qy	326	ACCTTGGAGAGACG	CTGTCAAGAGATGG	CGCTCAAGCAGGAT	TGTGATCCAGCACGCC
Db	1384	CGTGCGGGGAACG	CGCCACGAAGATCG	GGCTTCGGCATCT	CTGCTCAGCAANGAGCCC
Qy	386	GAGTGGTGGAGTCC	ACCCGGGACGAATT	TGACTTNCCCAACA	CAACTGGCC 434
Db	1444	CGGTGTGGCCCTAC	AACCGGGGAGCAGC	CCCCATCTTCGTCA	AACTCCCC 1492

RESULT 13
US-08-944-496-36
; Sequence 36, Application US/08944496
; Patent No. 6124433
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York


```

: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: NY
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSEQ Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/826,246
: FILING DATE: 28-MAR-1997
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/799,910
: FILING DATE: 13-FEB-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/011,787
: FILING DATE: 16-FEB-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Coruzzi, Laura A
: REGISTRATION NUMBER: 30,742
: REFERENCE/DOCKET NUMBER: 7853-078-999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212)7909090
: TELEFAX: (212)8699741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3084 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: cdna
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 1032...1736
: OTHER INFORMATION:
:
: US-08-826-246-11

```

```

Query Match      7.08; Score 41.6; DB 3; Length 3084;
Best Local Similarity 52.7%; Pred. No. 0.086;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 266 ACGCCCATCTGGTCAAGGGCCACGGCTGGACACCGCCACTTTGACAAGATCAAGCAGT 325
    || || || || || || || || || || || || || || || || || || || ||
1324 ACCTACCTCAGGCGCGGCTTCTGCTGGGCCAGCTCAACCTGGAGCAGCGCAGCAGT 1383

Qy 326 ACCTTGGAGAGCGCTGCAAGAGATGGGCTCAAGCAGGATGTGATCCAGCAGCGCGCGG 385
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1384 CGTGGCGGACGCGCAGCAGATCGGCTTCGGCATCTGCTCAGCAGGAGCGCGACG 1443

Qy 386 GAGTGGTGGAGTCCACCGGACGAAATTGACTTCCCAACAACTGCGG 434
    |||| || || || || || || || || || || || || || || || || || ||
Db 1444 GCCTGGGCTTACAACCGCGGAGCAGCCCATCTTCGTCAACTCCCC 1492

```

Search completed: April 23, 2003, 13:50:28
Job time : 58 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2003, 12:35:20 ; Search time 103 seconds
(without alignments)
6317.485 Million cell updates/sec

Title: US-09-920-953-2

Perfect score: 598

Sequence: 1 gcctgcctgcagacgcggc.....agccattttgcaccagcc 598

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:

- 1: /cgn2_6/ptodata/2/pubnpa/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubnpa/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubnpa/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubnpa/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubnpa/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubnpa/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubnpa/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubnpa/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubnpa/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubnpa/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubnpa/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubnpa/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubnpa/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubnpa/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	7.0	1058	10	US-09-452-239-11
2	41.6	7.0	1280	10	US-09-923-922-4
3	41.6	7.0	1280	10	US-09-934-456-1590
4	41.6	7.0	1817	10	US-09-924-417-64
5	41.6	7.0	3083	9	US-09-970-820-36
6	41.6	7.0	3083	9	US-09-985-718-36
7	41.6	7.0	3083	10	US-09-371-900-36
8	41.6	7.0	3083	10	US-09-924-417-62
9	41.6	7.0	3084	8	US-08-825-486-11
10	41.6	7.0	3084	8	US-08-870-434-6
11	41.6	7.0	3084	9	US-09-560-150-6
12	41.6	7.0	3084	10	US-09-372-044-11
13	41	6.9	353	10	US-09-864-761-17688
14	41	6.9	355	10	US-09-864-761-25057
15	41	6.9	582	10	US-09-864-761-8326
16	39.4	6.6	417	9	US-09-918-995-3955
17	38.2	6.4	1341	9	US-09-975-719-133
18	38.2	6.4	2176	8	US-08-808-031A-1
19	38	6.4	1566	10	US-09-815-242-7758

Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence 1, Appli	Sequence
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	----------

ALIGNMENTS

RESULT 1

US-09-452-239-11
; Sequence 11, Application US/09452239
; Patent No. US20020081693A1
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Pader, Gary M.
; APPLICANT: Cahoon, Rebecca E.
; TITLE OF INVENTION: Plant Caffeyol-coA O-Methyltransferase
; FILE REFERENCE: BB1284 US NA
; CURRENT APPLICATION NUMBER: US/09/452,239
; CURRENT FILING DATE: 1999-12-01
; EARLIER APPLICATION NUMBER: 60/110,594
; EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 11
; LENGTH: 1058
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-452-239-11

Query Match 7.0%; Score 42; DB 10; Length 1058;

Best Local Similarity 52.9%; Pred No. 0.0021;

Matches 90; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 162 CGAGTCCCTGGACATGCAAGAGCAGCAGAGTCAAGTTCATGAGCTTCGTGT 221

Db 619 CAACFACCTCAACTACCAGCGGCTGATGAAGTTCGAGGCGGCGCTCGT 678

QY 222 TGGCGGAGCAGACCAATACAGGCGGAGGAGCATGTACGACGACAGCCCATCTGTC 281

Db 679 CTACGACACACAGCTGTGAGACGGCTCGTGTCTCCCGCGGAGCCCATGCGCAA 738

QY 282 GGGCCACGGCGCTGGACCGCCACTTTGACAAGATCAAGCAGTACTCTTG 331

Db 739 GTACATCCGCTACTACCGGCACTTCTGCTCGAGCTCAACAAGGCCCTCG 788

RESULT 2

US-09-923-922-4

; Sequence 4, Application US/09923922

Patent No. US20020106732A1
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohei
APPLICANT: Kawabata, Masahiro
TITLE OF INVENTION: SHAD6 AND USES THEREOF
FILE REFERENCE: L0461/7120
CURRENT APPLICATION NUMBER: US/09/923,922
PRIOR FILING DATE: 2001-08-07
PRIOR APPLICATION NUMBER: 09/096,776
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: US 60/049,990
PRIOR FILING DATE: 1997-06-13
PRIOR APPLICATION NUMBER: US 60/053,040
PRIOR FILING DATE: 1997-07-18
PRIOR APPLICATION NUMBER: US 60/066,173
PRIOR FILING DATE: 1997-11-18
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 1280
TYPE: DNA
ORGANISM: Homo sapiens
S-09-923-922-4

Query Match 7.0%; Score 41.6; DB 10; Length 1280;
Best Local Similarity 52.7%; Pred. No. 0.003;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 266 ACCCCCATCTGGTCAAGGCGCAGCGCTGGACACCGCCACTTTTGACAAGATCAAGCAGT 325
DB 405 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCGCAGCTCAACCTGGAGCGCAGCGAGT 464
QY 326 ACCTTGGAGAGCGCTGCAAGAGATGGCGCTCAAGCAGGATGTGATCCAGCAGCGCGC 385
DB 465 CGGTGGCGGCGACGCGCAGCAGATCGGCTTCGGCATCTCTGCTCAGCAAGAGCGCGAGC 524
QY 386 GAGTGGTGGAGTCCACCGCGCAGCAATTTGACTTNCCTCAACAACTGCGC 434
DB 525 CGGTGGCGGCTTACAACCGCGCGAGCAGCCCATCTTCTGTCAACTCCCC 573

RESULT 3
US-09-954-456-1590
Sequence 1590, Application US/09954456
Patent No. US20020115057A1
GENERAL INFORMATION:
APPLICANT: Young, Paul
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
FILE REFERENCE: 689290-76
CURRENT APPLICATION NUMBER: US/09/954,456
PRIOR FILING DATE: 2001-09-18
PRIOR APPLICATION NUMBER: US/60/233,617
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,052
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,923
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,134
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,637
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,638
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,711
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,720
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,840
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,863
PRIOR FILING DATE: 2000-09-27
NUMBER OF SEQ ID NOS: 2276

SOFTWARE: PatentIn version 3.0
SEQ ID NO 1590
LENGTH: 1280
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-456-1590
Query Match 7.0%; Score 41.6; DB 10; Length 1280;
Best Local Similarity 52.7%; Pred. No. 0.003;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 266 ACCCCCATCTGGTCAAGGCGCAGCGCTGGACACCGCCACTTTTGACAAGATCAAGCAGT 325
DB 405 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCGCAGCTCAACCTGGAGCGCAGCGAGT 464
QY 326 ACCTTGGAGAGCGCTGCAAGAGATGGCGCTCAAGCAGGATGTGATCCAGCAGCGCGC 385
DB 465 CGGTGGCGGCGACGCGCAGCAGATCGGCTTCGGCATCTCTGCTCAGCAAGAGCGCGAGC 524
QY 386 GAGTGGTGGAGTCCACCGCGCAGCAATTTGACTTNCCTCAACAACTGCGC 434
DB 525 CGGTGGCGGCTTACAACCGCGCGAGCAGCCCATCTTCTGTCAACTCCCC 573

RESULT 4
US-09-924-417-64
Sequence 64, Application US/09924417
Patent No. US20020142441A1
GENERAL INFORMATION:
APPLICANT: Falb, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/924,417
FILING DATE: 07-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/034,286
FILING DATE: 04-MAR-1998
APPLICATION NUMBER: 08/870,434
FILING DATE: 06-JUN-1997
APPLICATION NUMBER: 08/799,910
FILING DATE: 13-FEB-1997
APPLICATION NUMBER: 60/011,787
FILING DATE: 16-FEB-1996
APPLICATION NUMBER: 08/599,654
FILING DATE: 09-FEB-1996
APPLICATION NUMBER: 08/485,573
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: 08/386,844
FILING DATE: 10-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-114-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE

```

; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1817 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 64:
US-09-924-417-64

Query Match          7.0%; Score 41.6; DB 10; Length 1817;
Best Local Similarity 52.7%; Pred. No. 0.0034;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACGCCCATCTGTCAGAGCCACGGCTCGACACCGCCACCTTTGACACAGATCAAGCAGT 325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1230 ACCTACCTCAGGCGACGGCTTCTGCTGGCCAGCTCAACCTGGAGCAGCGCAGGAGT 1289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 326 ACCTTGGAGAGCGTGTCAAGAGATGGCGGTCAAGCAGGATGTGATCCAGCAGCGCGCG 385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1290 CGGTGGCGGCAACGCCACAGATCGGCTTCGGCATCTCTGCTCAGCAAGGAGCCGACG 1349
QY 386 GAGTGTGAGTCCACCGCGCAGCAATTTGACTTNCACCAACTGCCG 434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1350 GCGTGTGGGCTTACAAACCGCGGAGCACCCTCATCTTCGTCAACTCCCC 1398

RESULT 5
US-09-970-820-36
; Sequence 36, Application US/09970820
; Patent No. US20020170077A1
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/970,820
; FILING DATE: 05-OCT-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16

```

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-970-820-36

Query Match          7.0%; Score 41.6; DB 9; Length 3083;
Best Local Similarity 52.7%; Pred. No. 0.0041;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 AGCCCATCTGTCAGAGCCACGGCTCGACACCGCCACCTTTGACACAGATCAAGCAGT 325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1324 ACCTACCTCAGGCGACGGCTTCTGCTGGCCAGCTCAACCTGGAGCAGCGCAGGAGT 1383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 326 ACCTTGGAGAGCGTGTCAAGAGATGGCGGTCAAGCAGGATGTGATCCAGCAGCGCGCG 434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1384 CGGTGGCGGCAACGCCACAGATCGGCTTCGGCATCTCTGCTCAGCAAGGAGCCGACG 1443
QY 386 GAGTGTGAGTCCACCGCGCAGCAATTTGACTTNCACCAACTGCCG 434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1444 GCGTGTGGGCTTACAAACCGCGGAGCACCCTCATCTTCGTCAACTCCCC 1492

RESULT 6
US-09-986-718-36
; Sequence 36, Application US/09986718
; Patent No. US20020178458A1
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/986,718
; FILING DATE: 09-NOV-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/485,573
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16

```

```

; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-986-718-36

Query Match      7.0%; Score 41.6; DB 9; Length 3083;
Best Local Similarity 52.7%; Pred. No. 0.0041;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACGCCCTCTGGTCAAGGCCACGCCCTGGGACCCGCTTTGACAAAGATCAAGCAGT 325
      || || || || || || || || || || || || || || || || || || || ||
Db 1324 ACCTACCTTCAGGGCAGCGGCTTCTGCTTGGGCCAGCTCAACCTGGAGCAGCGCAGT 1383
      || || || || || || || || || || || || || || || || || || || ||
QY 326 ACCTTGGAGAGACGCTGCAAGAGATGGCGCTCAAGCAGGATGTGATCCAGCAGCGCGCG 385
      || || || || || || || || || || || || || || || || || || || ||
Db 1384 CGGTGCGGGAACGGCGAGCAGATCGGCTTGGGCATCTGCTAGCAAGAGCGCCGACG 1443
      || || || || || || || || || || || || || || || || || || || ||
      386 GAGTGGTGGAGTCCACCGCGGACGAAATTTGACTTNCCTCAACAACTGCCG 434
      || || || || || || || || || || || || || || || || || || || ||
Db 1444 CGGTGTGGGCTTACAACCGCGGAGCACCCTATCTTCTGTCAACTCCCC 1492
      || || || || || || || || || || || || || || || || || || || ||

RESULT 7
US-09-371-900-36
; Sequence 36, Application US/09371900
; Patent No. US20020137700A1
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/371,900
; FILING DATE: 11-Aug-1999
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
```

```

; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-371-900-36

Query Match      7.0%; Score 41.6; DB 10; Length 3083;
Best Local Similarity 52.7%; Pred. No. 0.0041;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACGCCCTCTGGTCAAGGCCACGCCCTGGGACCCGCTTTGACAAAGATCAAGCAGT 325
      || || || || || || || || || || || || || || || || || || || ||
Db 1324 ACCTACCTTCAGGGCAGCGGCTTCTGCTTGGGCCAGCTCAACCTGGAGCAGCGCAGT 1383
      || || || || || || || || || || || || || || || || || || || ||
QY 326 ACCTTGGAGAGACGCTGCAAGAGATGGCGCTCAAGCAGGATGTGATCCAGCAGCGCGCG 385
      || || || || || || || || || || || || || || || || || || || ||
Db 1384 CGGTGCGGGAACGGCGAGCAGATCGGCTTGGGCATCTGCTAGCAAGAGCGCCGACG 1443
      || || || || || || || || || || || || || || || || || || || ||
      386 GAGTGGTGGAGTCCACCGCGGACGAAATTTGACTTNCCTCAACAACTGCCG 434
      || || || || || || || || || || || || || || || || || || || ||
Db 1444 CGGTGTGGGCTTACAACCGCGGAGCACCCTATCTTCTGTCAACTCCCC 1492
      || || || || || || || || || || || || || || || || || || || ||

RESULT 8
US-09-924-417-62
; Sequence 62, Application US/09924417
; Patent No. US20020142441A1
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; THE TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR
; DISEASE
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/924,417
; FILING DATE: 07-Aug-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,286
; FILING DATE: 04-MAR-1998
; APPLICATION NUMBER: 08/870,434
; FILING DATE: 06-JUN-1997
; APPLICATION NUMBER: 08/799,910
; FILING DATE: 13-FEB-1997
; APPLICATION NUMBER: 60/011,787
; FILING DATE: 16-FEB-1996
; APPLICATION NUMBER: 08/599,654
; FILING DATE: 09-FEB-1996
; APPLICATION NUMBER: 08/485,573
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-114-999
```

; TYPE: nucleic acid .

Query Match	7.0%;	Score 41.6;	DB 8;	Length 3084;
Best Local Similarity	52.7%;	Pred. No. 0.0041;		

```
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 266 AGCCCATCTGTCAAGGCCACGGCTGGACACCGCCACTTTGACAAGATCAAGCAGT 325
  || || || || || || || || || || || || || || || || || || || ||
Db 1324 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACTGGAGCAGCGCAGGT 1383
QY 326 ACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCGCGG 385
  || || || || || || || || || || || || || || || || || || || ||
Db 1384 CGGTGGGGGAACGGCGACCAAGATCGGCTTCGGCATCTGCTCAGCAAGAGCGCGAG 1443
QY 386 GAGTGGTGAGTCCACCGCGACGAATTTGACTTNCACCAACTGCGC 434
  || || || || || || || || || || || || || || || || || || || ||
Db 1444 GCGTGGGCTTACAAACCGCGGAGCACCCCATCTTGTCAACTCCCC 1492

RESULT 11
US-09-560-150-6
; Sequence 6, Application US/09560150
; Publication No. US20030073076A1
; GENERAL INFORMATION:
; APPLICANT: FALB, Dean A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 7853-126
; CURRENT APPLICATION NUMBER: US/09/560,150
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 09/126,640
; PRIOR FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: 08/870,434
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 08/799,910
; PRIOR FILING DATE: 1997-02-13
; PRIOR APPLICATION NUMBER: 60/011,787
; PRIOR FILING DATE: 1996-02-16
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3084
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-560-150-6

Query Match
Best Local Similarity 7.0%; Score 41.6; DB 9; Length 3084;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 266 ACGCCCATCTGTCAAGGCCACGGCTGGACACCGCCACTTTGACAAGATCAAGCAGT 325
  || || || || || || || || || || || || || || || || || || || ||
Db 1324 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACTGGAGCAGCGCAGGT 1383
QY 326 ACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCGCGG 385
  || || || || || || || || || || || || || || || || || || || ||
Db 1384 CGGTGGGGGAACGGCGACCAAGATCGGCTTCGGCATCTGCTCAGCAAGAGCGCGAG 1443
QY 386 GAGTGGTGAGTCCACCGCGACGAATTTGACTTNCACCAACTGCGC 434
  || || || || || || || || || || || || || || || || || || || ||
Db 1444 GCGTGGGCTTACAAACCGCGGAGCACCCCATCTTGTCAACTCCCC 1492

RESULT 12
US-09-372-044-11
; Sequence 11, Application US/09372044A
; Patent No. US20020102603A1
; GENERAL INFORMATION:
; APPLICANT: Dean FALB et al.
; TITLE OF INVENTION: Compositions and Methods for the
; FILE REFERENCE: 7853-152
; CURRENT APPLICATION NUMBER: US/09/372,044A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
```

```
LENGTH: 3084
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1032)...(1736)
US-09-372-044-11

Query Match
Best Local Similarity 7.0%; Score 41.6; DB 10; Length 3084;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 266 AGCCCATCTGTGTCAAGGCCACGGCTGGACACCGCCACTTTGACAAGATCAAGCAGT 325
  || || || || || || || || || || || || || || || || || || || ||
Db 1324 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACTGGAGCAGCGCAGGT 1383
QY 326 ACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCGCGCG 385
  || || || || || || || || || || || || || || || || || || || ||
Db 1384 CGGTGGCGGAACGGCGAGCAAGATCGGCTTCGGCATCTGCTCAGCAAGAGCGCGAG 1443
QY 386 GAGTGGTGAGTCCACCGCGACGAATTTGACTTNCACCAACTGCGC 434
  || || || || || || || || || || || || || || || || || || || ||
Db 1444 GCGTGGGCTTACAAACCGCGGAGCACCCCATCTTGTCAACTCCCC 1492

RESULT 13
US-09-864-761-17688/C
; Sequence 17688, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
```


XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 PR 29-OCT-1999; 99US-0162143.

Alignment Scores:

Pred. No.: 0.000838 Length: 129
 Score: 118.00 Matches: 43
 Percent Similarity: 45.9% Conservative: 19
 Best Local Similarity: 31.8% Mismatches: 37
 Query Match: 11.0% Indels: 36
 DB: 21 Gaps: 11

US-09-920-953-2 (1-598) x AAG12569 (1-129)

QY 80 TGGGGCGCCAGACGATGAGCTGGCGG-----TTGACACCTTCT 121
 1 TTPAlaAlaArg-----TTPArgHisCysAlaGlyArgProThrAlaProThr 16
 QY 122 AGGATAAAGTGTGCTGAC-----CGAGCTGTGCTG----- 154
 17 SerThrArgCysTrpAlaSerIleCysSerAlaArgProCysCysAlaSerArgAla 36
 QY 155 -----CCTCTTCAGTCC-----TGGACATGC---AAGACGAGA 187
 37 PheArgTrpProSerSerThrProSerArgSerArgSerTrpCysCysTrpGlySerArg 56
 QY 188 ACATGACGAGCTCAAGTTCATGCTTGTGTTGGCGGACGACGACCAATACAGGGCC 247
 57 ArgArgSerProAlaProArgProSerSerCysThrAla-----SerSerArgArgSer 74
 QY 248 GAAGCATGTACAGCAGACGCCATCTGCTCAAGGGCCAGCGCTGGACCCGCGCACT 307
 75 SerArgThrProSerThrSerPro-----SerArgSerSerCysArgArgArgAlaSer 92
 QY 308 TTGACAGATCA---AGCAGTACCTTGGAGACAGCTGCAAGAGATGGCGCTCAAGCAGG 364
 93 CysSerArgAlaProThrSerProProSerArgTrpCysCysThrTrpTrpSerAlaGly 112
 QY 365 ATGTGATCCAGCAGCCCGCGGAGTGG-----TGGAGTCCACCC 403

Db 113 Ser---SerSerThrArgSerAlaTrpAlaCysTrpGlyProPro 126
 RESULT 2
 AAB95825
 ID AAB95825 standard; Protein; 257 AA.
 XX AAB95825;
 AC AAB95825;
 XX 26-JUN-2001 (first entry)
 XX Human protein sequence SEQ ID NO:18834.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW Homo sapiens.
 OS EP1074617-A2.
 PN EP1074617-A2.
 XX 07-FEB-2001.
 PD 28-JUL-2000; 2000EP-0116126.
 PF 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -
 Claim 8; SEQ ID 18834; 2537pp + CD ROM; English.
 The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

SQ Sequence 257 AA;

Alignment Scores:
 Pred. No.: 0.00139 Length: 257
 Score: 117.00 Matches: 60
 Percent Similarity: 38.10% Conservative: 20
 Best Local Similarity: 28.57% Mismatches: 72

```
Query Match: 10.97% Indels: 58
DB: 22 Gaps: 13

US-09-920-953-2 (1-598) x AAB95825 (1-257)

QY 26 GCACACAGACAGCGCGGATCGCGGCGCAAGAGCTGTTTATGACCTGGCGC 85
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 AlaProSerLysAsnSerSerArg---LeuGlyGlyArgProCysMetCysThr---Ala 51
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 86 GCGAGAAGGCGATGAAGTGGCGGTTCACACCTCTACGATAAGGTGCTGACCGCG 145
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52 GlyArgArg-----ProAsnArgAlaSerGlyArgArgArg 64
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 146 ACCTGCTCCCTCTCTTCAGT-----CCCTGGACATGC-----AAGAGCAGAGATGA 193
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 SerCysSerProAlaProThrTriProProLeuCysCysTyrProGlnSerArgProThr 84
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 194 ACCAGCTCAAGTTCATGACCTTCGTGTTGGCGGAGC-----AGACCA----- 236
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 AlaSerAlaAlaGlyProGlyAlaCysMet-ArgAlaSerGlyArgProHisGlyAsnTh 104
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 237 -----ATACAGGCGCGAAGCATGTACGAGCGCACGCGCATCTGCTCAAGGG-- 284
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 rThrAlaSerThrAlaProProArgHisProArgProArgProGlyGlyProAlaLe 124
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 285 -----CCACGG-----CCTGGACCCCGCCACTTTGACAAAGAT----- 317
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 uArgProThrProArgProCysAlaGlyProAlaProProProAlaSerArgAspCysAr 144
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 318 -----CAACGAGTACCTTCGAGAGACCGCTGCAAGACATGGGCTCAAGCAGGA---TG 367
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 gCysArgArgProArgTgtTProArgAlaGlyArgArgGlyArgArgAlaGlyAlaCy 164
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 368 TGATCCAGCAGCGCGCGAGTGGTGGAGTCCACCGCGGACGAATTCGCTTCCCAACA 427
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 sLysProSer-CysAlaGlyAlaAlaTrpSerAlaArgGly----- 177
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 428 ACTGCGCACCACTGATTTTCATTAACCCCAACCCCGCTGAGCGCTCATTCATCCCAT 487
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 -----AlaProLeuCysSerTyrArgThrS 186
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 488 TTTGAGCGGGAGCGCCAGTTGCGCGAGCGCGCCCGCGCGGCGCCAGGAGCTTGAATCG- 546
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 186 eCysAlaGlySerCysGlyAlaArgThrAlaProThrProAlaProThrCysAlaSerP 206
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 547 -----TTTCCAGCGCTTGTCTGC 564
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 206 rOSerAlaAlaAlaSerSerCysCys 214
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

--SULT 3
AAU54642
ID AAU54642 standard; Protein: 290 AA.
XX
AC AAU54642;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #15538.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI: 2001-616774/71.
DR N-PSDB: AAS95566.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
PS Example 1; SEQ ID No 15837; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 290 AA:

Alignment Scores:
Pred. No.: 0.00185 Length: 290
Score: 116.00 Matches: 44
Percent Similarity: 35.11% Conservative: 22
Best Local Similarity: 23.40% Mismatches: 74
Query Match: 10.87% Indels: 48
DB: 22 Gaps: 5

US-09-920-953-2 (1-598) x AAU54642 (1-290)
QY 17 CGGCCACAGCACACAGACGCGGAGCGGATCGCGGCGCAAGAGCTGTTTGTATG 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15 ArgArgProLeuSerCysArgSerArgSerGluThrThrAlaArgAlaCysMetPhe 34
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 77 ACCTGGCGCGCGCAGAAGGCA----- 97
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35 ThrLeuArgSerGlyArgAlaAlaAsnLeuLysSerLeuThrSerProAspThrProAsn 54
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 98 ---TGAGCTGGCGGTGACACCTTCTACGATAGGTGCTGCGGAGCGGAGCTGCTGC 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 SerLeuThrTrpProAlaThrThrSerAlaAlaSerSerThrHisArgCysTrp 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 155 CCTTCTTCGAGTCCCTCGACATGCAAGCAGCAGAGATGAAGCAGGTCAAGTTCATGAGCT 214
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 ProSerLeuThrProArgProThrArgThrThrValTrpCysArgAlaSerArgProGln 94
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 215 TCGTGTGTCGGGAGCAGACCAATACAGGCCCGCAAGCATGTACGACGACACCCCATC 274
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 SerIleTrpCysIleHisSerAlaThrAlaProProAlaCysValSerArgSerProAla 114
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 275 TGGTCAAGGCGCCAGCGCC----- 292
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 GlnIleArgArgProSerAlaSerAsnSerValAlaGlnIleArgLeuProThrLeuThr 134
    ||||| :|: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

QY 293 -----TGGACACCGCCACTTTCACAGATCAACGACTACCTTG 331
Db 135 TrpProLeuProCysTyrTrpProAlaTrpThrAlaSerArgThrArgSerArg 154
QY 332 GAGACAGCTGCAAGAGATGGCGCTCAAGCAGATGTGATCCAGCAGCGCGGAGTGG 391
Db 155 ProArgSerThrArgThrSerThrSerLeuLeuArgSerWetProPro---Trp 173
QY 392 -----TGG-----AGTCCACCGCGGAGCAATTG 415
Db 174 IleThrCysArgAlaThrTrpValArgSerThrIleSerArgProThrThrSerPhe 193
QY 416 ACTTNCACCAACTCGCACCA 439
Db 194 SerLeuProValThrCysSerPro 201

RESULT 4
ID ABG03530 standard; Protein; 599 AA.
X ABG03530;
DT 13-FEB-2002 (first entry)
XX Novel human diagnostic protein #3521.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX N-PSDB: AAS67717.
XX New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity -
Claim 20; SEQ ID No 3389; 103pp; English.
The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG00010-ABG30377 represent novel human
diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 599 AA;
Alignment Scores:
Score: 0.00352 Length: 599
Pred. No.: 114.50 Matches: 58
Percent Similarity: 33.18% Conservative: 14
Best Local Similarity: 26.73% Mismatches: 92
Query Match: 10.20% Indels: 53
DB: 22 Gaps: 12

US-09-920-953-2 (1-598) x ABG03530 (1-599)

QY 583 TGGCTGATGGCTCTTCAATGCAAGAGGGCTGG-----CAACGATTTCAGGCTC 533
Db 74 TrpThrProHisCysProAlaSerAlaAlaTrpGlyAlaGluGlnGluProValArgSer 93
QY 532 CTGGGCCCTCTGGCGCGCTC-----GGCACTGGCGCTCCCGCTCAAAATC 485
Db 94 TrpGlyProArgAlaSerGlnSerHisCysProGlyGlyLeuArgAlaProProGly 113
QY 484 GAT---GGAATGAGCGCTCAGCTGGGTTGGTTAATGAAATCATCTGGTGGCGAGT 428
Db 114 SerValArgCysSerThrGln***AspCysSerSerValArgProAlaTrpSerArgSer 133
QY 427 -----TGTTGGNAAAGTCAAAATTCGTCGGCGGTGG 398
Db 134 **GlyAlaCys**GlnVal***ProArgCysProCysArgThrProAlaThrGlyTrp 153
QY 397 ACTCCACCACCTCGCGCGGTGCTGGA-----TCACATCTGCT 359
Db 154 AlaProProGlnGlyArgCysGlyProThrThrAlaProGlySerThrGlyProAla 173
QY 358 TGACGCCCATCTCTTCAGCGCTCTCTCAAGGTACTGCTTGATCTTGTCAAAGTGGCGGT 299
Db 174 GlyArgAlaSerLeuCysCysProArgArgAlaHisLeu----- 186
QY 298 GGTCCAGCGCGTGGCGCTTGACCAGATGGCGGTGTCGTACATGCTTCGGCCCT--- 242
Db 187 ---ProGly***TrpPro-----GlnLysLeuLysCysAlaHisProGly 200
QY 241 -----TGATTGGTCTGCTCCGCCAAACACAGCTCATGAATGCTGCTGCTCATCT 188
Db 201 AlaLysSerLeuGlyLeuAlaCysGlnProHisArgGlyLysGlyThrProLleGluGly 220
QY 187 TCTGCTCTTTCATGCTCCAGGACTCGAAGAGGCGACGACCTCCGGGTGACCGACCT 128
Db 221 ProAlaCysGlyThr***GlyGlyArgArgGlySerGlyCysProGlyArgProHisThr 240
QY 127 TATCGTAGAAGGTGTCAACCGCGCTTCATGCTCTTCGCGCGCCAGGTCTATCAACA 68
Db 241 ---ArgArgArgCys***ProProAlaProCys-----GlyArgArgSer 254
QY 67 GCTTCTTGGCGCGCATCCCGCTTCGCGCTCTCTGCTGCTGCTGGTGGCGCG 17
Db 255 AlaGlySerAlaHis---ProAlaArgPro-----TrpPro 265

RESULT 5
AAG12569
ID AAG12569 standard; Protein; 129 AA.
XX AAG12569;
XX 17-OCT-2000 (first entry)
XX Zeta mays protein fragment SEQ ID NO: 11732.
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.

PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0156659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 18-OCT-1999; 99US-0159638.
 PR 21-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 23-OCT-1999; 99US-0161404.
 PR 23-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
 Pred. No.: 0.00412
 Score: 111.50
 Percent Similarity: 38.85%
 Test Local Similarity: 28.78%
 Very Match: 9.94%
 JB: 21

US-09-920-953-2 (1-598) x AAG12569 (1-129)

QY 410 TCGTCGGGCTGACCTCCACCATCCGGCGGCTGGATCATCTCTGACGCC 351
 Db 9 AlaGlyArgProThrAlaProThrSerThrArgCysTrpAlaSerIleCys----- 25
 QY 350 ATCTCTGACGCTCTCCAGGTACTGCTTGTCTGCTTGTCAAAG---TGGCGGTGGTCC 294
 Db 26 -----SerAlaArgProCysCysAlaSerArgAlaPheArgTrp-Pr 40
 QY 293 AGCCGCTGGCCCTTGACCATGGCGTGGCTGTACATGCTTCGGCCCTTGATGG 234
 Db 40 oSerSerThrPro-----SerArgSerArgSerTrpCys----- 51
 QY 233 TCTGCTCCGCCAAACACAGACATCACTGACCTGCTTCATCTCTCTCTGTCATG 174
 Db 52 -----Cy 52
 QY 173 TCCAGGACTCGAAGAGGCGACAGCTCGGGTCAAGCCA-----GCACCTTA 126
 Db 52 stPGlySerArgArgSerProAlaPro---ArgProSerSerCysThrAlaSer 71

QY 125 TCGTAGAAGGTGTCACCGCCAGCTTCATGCTTCTGCGCGCCAGGTCATCAACAGC 66
 Db 71 rArgArgSerSerArgThrProSerThrSerProSerArgSerSerCysArgArgAl 91
 QY 65 TTCTTGGCGCCGATCCCGCTT---CGCGCGTCTCTGTGTGTCG-----TGGTGG 20
 Db 91 aserCysSerArgAlaProThrSerProSerArgTrpCysCysThrTrpTrp 109
 RESULT 6
 AAW18663
 ID AAW18663 standard; Protein; 387 AA.
 XX
 AC AAW18663;
 XX
 DT 24-JUL-1997 (first entry)
 XX
 DE Fragmented human NF-H gene +2 frameshift mutant product.
 XX
 KW Frameshift mutation product; GAGA motif; somatic mutation; diagnosis;
 KW detection; antibody; probe; cancer; neoplasia; neurodegenerative;
 KW Parkinson's; Alzheimer's disease; Pick's; Huntington's disease;
 KW Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;
 KW PSP; amyotrophic lateral sclerosis; multiple sclerosis; MS;
 KW cardiovascular; rheumatoid arthritis.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1..387 /note= "X corresponds to a stop codon in the
 FT accompanying DNA file, AAT69796"
 FT
 XX WO9712992-A2.
 XX
 PD 10-APR-1997.
 XX
 PF 02-OCT-1996; 96WO-1B01106.
 XX
 PR 11-JAN-1996; 96US-0009832.
 PR 02-OCT-1995; 95GB-0020080.
 XX
 PA (ROVA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
 PA (UYRO-) UNIV ROTTERDAM ERASMUS.
 PA (UYUT-) UNIV STATE UTRECHT.
 XX
 PI Burchach JPH, Grosveid FG, Van Leeuwen FW;
 XX
 WPI: 1997-226235/20.
 DR N-PSDB; AAT69795.
 DR
 XX
 PT Use of mutant genes having frame:shift mutation(s) - for developing
 PT prods. for the diagnosis, prevention and treatment of associated
 PT diseases, e.g. cancer or neurodegenerative disease
 XX
 PS Claim 22; Fig 9; 123pp; English.
 XX
 CC AAW18663 and AAW18664 are +2 and +1 frameshift mutations, respectively,
 CC of a sequence comprising fragments of the coding sequence of the
 CC human neurofilament subunit NF-H gene corresponding to nucleotides
 CC 1-1162 of the wild-type NF-H gene. This region contains GAGAC motifs.
 CC Frameshift mutants of the tau, ubiquitin, apolipoprotein E,
 CC microtubule-associated protein 2 (MAP-2), neurofilament subunit L, M
 CC and H and amyloid A4 genes are claimed. All these genes share a common
 CC GAGAN motif (N-A, G, C or T), which is the site of common GA
 CC dinucleotide deletion(s) that cause neurodegenerative disorders.
 CC Antigenic peptides used for the production of antibodies, and small
 CC nucleic acid sequences derived from frameshift mutants are used in the
 CC diagnosis, prevention and treatment of cancer and neurodegenerative
 CC diseases, e.g. Parkinson's disease, Alzheimer's disease, Down's
 CC syndrome, frontal lobe dementia (Pick's disease), progressive
 CC supranuclear palsy (PSP), amyotrophic lateral sclerosis, Huntington's
 CC disease, multiple sclerosis, and other degenerative diseases such as

CC cardiovascular disease and rheumatoid arthritis.

XX Sequence 387 AA;

Alignment Scores:

Pred. No.: 0.00622 Length: 387
Score: 111.50 Matches: 40
Percent Similarity: 35.95% Conservative: 15
Best Local Similarity: 26.14% Mismatches: 49
Query Match: 10.45% Indels: 49
DB: 18 Gaps: 7

US-09-920-953-2 (1-598) x AAH18663 (1-387)

Qy 11 CAGACGCGCCACACACAGACGCGGATGCGGCGGCAAGAGCTGT 70

Db 161 GlnAlaProGlnAlaProThrArgTrpThrArg***AlaThrGlyArgAla-- 179

Qy 71 TTGATGACCTGGCGCGCCAGAGCATGAAGCTGGCGGTGCACACCTTCTACGATAAGG 130

180 -----AlaTrp-----TTPArgTrpProHisAlaValArg 190

Qy 131 TGCTGGCTGACCCGAGCTGCTGC-----CCTTCTCGAGTCCC----- 169

Db 191 -----ArgSerCysArgArg***ThrThrAlaSerProGlyThrSerThr 206

Qy 170 -----TGCACATGCAAGCAGCAGATGAAGC 196

Db 207 ArgCysGlySerTrpArgArgThrThrAlaAlaTrpArgAlaArgLeuArgCysGly 226

Qy 197 AGGTCAAGTTCATGAGCTCGTGTGGCGGAGCAGCAATACAAAGCGCGAGCATGT 256

Db 227 SerSerArgArgAlaAlaProLeuTrpAlaSerCysThrSerAlaArgSerAlaArgCys 246

Qy 257 ACGAGCACACGCCATCTGTGCAAG-----GCCACGGCTGCACACCGCCACT 307

Db 247 AlaAlaArgCysCysAlaTrpAlaArgAlaValSerTyrAlaTrpSerArgSerThr 266

Qy 308 TTGACAAGTCAACAGCAGTACCTTG-----TTPArgTrpProGlySerGluArg 286

Db 267 CysSerArgThrSerArgThrCysAlaSerAla***ThrThrArgProGlySerGluArg 286

Qy 332 -----GAGACACGCTGCAAGATGGCGGTCACAGCAGG 364

Db 287 ArgProArgArgProAlaArgTrpArgAlaSerArg 299

RESULT 7

AY04954

AAI04954 standard; Protein; 572 AA.

AC AAY04954;

XX 06-JUL-1999 (first entry)

DE Mycobacterium species protein sequence 41T#2.

XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;

XX hybridisation; detection; vaccine; immunisation; infection.

OS Mycobacterium sp.

XX WO9909186-A2.

XX 25-FEB-1999.

XX 14-AUG-1998; 98WO-FR01813.

XX 11-SEP-1997; 97FR-0011325.

XX 14-AUG-1997; 97FR-0010404.

XX (INSP) INST PASTEUR.

XX Gicquel B, Lim EM, Pelicic V, Portnoi D, Goguet de la Salmoniere Y;

PI Guigueno A;

XX WPI; 1999-181045/15.

DR N-PSDB; AAX34206.

XX Mycobacterial DNA vectors containing reporter constructs - for
identifying coding or promoter sequences involved in
infection-associated protein expression

PS Claim 32; Fig 41T; 309pp; French.

XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
proteins from various Mycobacterium species microorganisms. The
encoding nucleotide sequences can be used as primers and probes for
methods for detecting and identifying mycobacteria, especially belonging
to the M. tuberculosis complex. The encoded proteins can be used in
vaccines for immunisation against a bacterial or viral infection.

XX Sequence 572 AA;

Alignment Scores:

Pred. No.: 0.00814 Length: 572
Score: 111.00 Matches: 58
Percent Similarity: 30.25% Conservative: 14
Best Local Similarity: 24.37% Mismatches: 87
Query Match: 9.89% Indels: 79
DB: 20 Gaps: 12

US-09-920-953-2 (1-598) x AAY04954 (1-572)

Qy 533 CTGGGCGCCCTGGGCGGCTCGGCACTGGCGCTCCCGCTCAAAATCGATGGAATGAG 474

Db 313 ProValProPro-----TTPArgTrpProGlyProAla 319

Qy 473 CGCTCAGCTGGGTTGGTTAATCAAAATCAGTTGGTGGCGCAGTTCTGGGNAAGTCA 414

Db 320 AspTrpArgTrpGlyGluLeuAlaGlySer***SerArgArgSerValProGlyProAla 339

Qy 413 AATTCGTCGCGGCTGGACTCCACCACCTCCGCGGCTGCTGCATCACATCTGCTGTGAGC 354

Db 340 AspCysArgProValAlaGlyArgGlyAlaAlaProCysTrpArgSerSerThrAlaThr 359

Qy 353 ---CCCATCTCTTCAGCGTC-----TCTCCAAAGTACTGC----- 321

Db 360 ValProProSerCysSerProGlyArgAlaProAlaCysCysAspArgValGlnThrPro 379

Qy 320 -----TTGATCTTGTCAAGTGG-----CGGTGCTCCAGCCGCGG 285

Db 380 ***HisArgProProLleSerValProThrSerTrpCysGlyProTrp***ThrProAla 399

Qy 284 CCCTTGACCATGGCGTGTGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 231

Db 400 ProArgThr***TrpCysCysProMetAlaMetTrpProProLysAsnTrpTrpProGly 419

Qy 230 GCTCCG-----TGTCCAGGACCTCGAAGAGGCGCAGCTCCGGGTCA 138

Db 420 ValProArgArgSerAlaGlyAlaSerThrTrpTrpProCysArgProAspArgTrpCys 439

Qy 224 -----CCAAACACGAAGCTCATGACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 176

Db 440 ArgGlyTrpProArgTrpProCysMet-ThrArgProAlaArgProSerThrThrAlaThr 459

Qy 175 -----TGTCCAGGACCTCGAAGAGGCGCAGCTCCGGGTCA 138

Db 459 rAlaTrpProValProProValLeuProGlyThrAspArgCysAlaLeuProProLysArg 479

Qy 137 G-----CCAGCACCTTATCGTAGAAGGTGTCACCCGCGCAGCTTCATCCCTCTCT 90

Db 479 gArg***ProGlyProValProAlaSerArg-----AlaThrValTrpVal 494

Qy 89 CGCGCGCCAGCTCATCAACAGCTTCTTGGCGCGCGCATCCCGCTTCCGCGCTCTCTGTG 30

Db 494 lSerArgAlaThrArgCys***SerSerProThrMetSerProArgProSerValTr 514

PF 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB; AA159947.
XX
Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries -
XX Example 2; SEQ ID NO 5722; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA138642-AA142213), with neurotropic,
CC immunosuppressant and cytotostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 317 AA;
XX
Alignment Scores:
Pred. No.: 0.0196 Length: 317
Score: 106.50 Matches: 64
Percent Similarity: 33.47% Conservative: 15
st Local Similarity: 27.12% Mismatches: 68
Very Match: 9.98% Indels: 89
DB: 22 Gaps: 12
XX
US-09-920-953-2 (1-598) x AA140791 (1-317)
QY 6 GCCTGCACACGGCCACACACACAGACGGCGGAGCGGATCGG----- 56
Db 62 AlaAlaGlnSerProHisArgGlyArgHisGlyGlyGlyAlaGlyLeuProPro 81
QY 57 -----GCCAAGAAGCTGTTGATGACCT----- 80
Db 82 ProArgSerProArgPheProGlnGlnSerVal-----ProAlaSerThrSerThrAla 99
QY 81 -----GGCGG 86
Db 100 ArgGlyProArgArgValSerArgLeuProProGlnHisProGlyProArgGlyArg 119
QY 87 CCCAAGACCGATGAAGCTGGCGGTGACACCTTACCATAGGTGCTGCGTACCGGA 146
Db 120 ArgArgArg-----ProGly 124
QY 147 GCTGCTGCCCTTCTTCGAGTCCCTGGACATGCAAGACGAGCAAGATGAAGCAGGTCAAGTT 206
Db 125 AlaGly-----ValGlyAlaProArgArgGlyArgAlaArgGlyGlnAlaGlyLeuLeu 142

QY 207 CATGAGCTTCGTGTTGGCGGAGCAGACCAATACAAGGGCGGAGCATGTACGACGCACA 266
Db 143 -----GlyArgGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlu 155
QY 267 CGCCCATCTGCTCAA-----GGGCCAGCGCTGGACACCGCCACTTTGACAAAGAT 317
Db 156 ArgAlaAlaLeuGlnAlaArgGlyArgGlyArgGlyProGlyProGluPro-----Asp 172
QY 318 CAAGCAGTA-----CCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGATGT 368
Db 173 GlnSerCysGlyArgProArgProArgAlaAlaAlaProGlyArg-----AlaProAla 191
QY 369 GATCCAGACCGCCCGGAGTGGTGGAGTCCACCGCAGCAATTTGACTTNCACCAACA 428
Db 192 AspPro--GlnProProAlaProArgProAlaProAlaProAlaProAlaProAla 211
QY 429 CTGGC-----CACCCAACTGATTTTCATTAAACCCCAACCCAGCGCTGAGC----- 472
Db 211 spAlaProAlaProAlaProAlaProAlaProProProProHisLeuGlyAlaLeuT 231
QY 473 -----GCTCATTCATCCATGATTTTGAGC 494
Db 231 hrAlaGlySerGlyGluGluArgGlnSerGlnProArgAlaGluThrLeuArgLeuGlyA 251
QY 495 GGGGAGCGCCAGTTCGCCAGCGCGCCAGCGGGGGCGCCAGGAGCCT 538
Db 251 rgGlyAlaProLeuProProArgAlaGluArgGlyGlyArgPro 265
RESULT 10
AA140792
ID AA140792 standard; Protein; 317 AA.
AC AA140792;
XX
XX 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 5723.
XX Human; neurotropic; immunosuppressant; cytotostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX N-PSDB; AA159948.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -
PS Example 2; SEQ ID NO 5723; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
X
J Sequence 317 AA;

Alignment Scores:

Pred. No.:	0.0196	Length:	317
Score:	106.50	Matches:	64
Percent Similarity:	33.47%	Conservative:	15
Best Local Similarity:	27.12%	Mismatches:	68
Query Match:	9.98%	Indels:	89
DB:	22	Gaps:	12

US-09-920-953-2 (1-598) x AAM40792 (1-317)	
QY 6 GCCTGCAGCGCGCCACACAGCAGCGCGGAGCGGGATGCGG-----	56
DB 62 AlaAlaGlnSerProHisArgArgGlyArgHisGlyGlyGlyAlaGlyLeuProPro 81	
QY 57 -----GCGCAAGAGCTGTTGTAGACCT-----	80
DB .82 ProArgSerProArgPheProGlnGlnUserVal-----ProAlaSerThrSerThra 99	
QY 81 -----GGCGCG 86	
DB 100 ArgGlyProArgValSerArgArgLeuProGlnHisProGlyProArgGlyArg 119	
QY 87 CGAGAGAGGATGAGCTGGCGGTGACACCTTCTAGATAGGTGCTGCTGACCGCGA 146	
DB 120 ArgArgArg-----ProGly 124	
QY 147 GCTGCTCCCTTCTTCAGTCCCTGGACATGCAAGACGACAGATGAGCAGGTCAAGTT 206	
DB 125 AlaGly-----ValGlyAlaProArgArgGlyArgAlaGlyGlnAlaGlyLeu 142	
QY 207 CATGAGCTTCTGCTTTGGCGGACGACCAATACAGGGCGGCGGATGTACGACGCACA 266	
DB 143 -----GlyArgGlnGlyGlnGlyGlyArgGlyAlaGluArgGlu 155	
QY 267 CGCCCATCTCGTCAA-----GGCCACGCGCTGGACCCACCGCATCTTGCACAGAT 317	
DB 156 ArgAlaAlaLeuGlnAlaArgArgGlyArgProGlyProGluPro-----Asp 172	
QY 318 CAAGCAGTA-----CCTTGGAGAGCGCTGCAAGAGATGGCGCTCAAGCAGGTGT 368	
DB 173 GlnSerCysGlyArgProArgArgAlaAlaAlaAlaProGlyArg-----AlaProAla 191	
QY 369 GATCCACGACCGCGCGCGCTGCTGGAGTCCACCGCGGCGGAAATTTGACTTNCACCAAA 428	
DB 192 AspPro--GlnProAlaProArgProAlaProAlaProAlaProAlaProAla 211	
QY 429 CTCGG-----CACCCAACTGATTTTCATTAAACCCACCGCGCTGAGC----- 472	
DB 211 spAlaProAlaProAlaProAlaProAlaProProProProHisLeuGlyAlaLeu 231	

QY 473 -----GCTCATTCATCATGATTTTTCGAC 494	
DB 231 hrAlaGlySerGlyGluArgGlnSerGlnProArgAlaGluThrLeuArgLeuGlyA 251	
QY 495 GGGAGCGCCAGTTGCGGAGCGCGCCAGCGGGCCAGGAGCCT 538	
DB 251 rgGlyAlaProLeuProProArgAlaGluArgGlyGlyArgPro 265	
RESULT 11	
AAU62666	
ID AAU62666 standard; Protein: 242 AA.	
XX	
AC AAU62666;	
XX	
DT 27-FEB-2002 (first entry)	
XX	
DE Propionibacterium acnes immunogenic protein #23562.	
XX	
KW SAPHO syndrome; synovitis; acne; pustulosis; hypotosis; osteomyelitis;	
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;	
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;	
KW dermatological; osteopathic; neuroprotectant.	
XX	
OS Propionibacterium acnes.	
XX	
PN WQ200181581-A2.	
XX	
PD 01-NOV-2001.	
XX	
PF 20-APR-2001; 2001WO-US12865.	
XX	
PR 21-APR-2000; 2000US-199047P.	
PR 02-JUN-2000; 2000US-208841P.	
PR 07-JUL-2000; 2000US-216747P.	
XX	
PA (CORI-) CORIXA CORP.	
XX	
PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;	
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;	
XX	
DR N-PSDB; AAS59628.	
XX	
PT Propionibacterium acnes polypeptides and nucleic acids useful for	
PT vaccinating against and diagnosing infections, especially useful for	
PT treating acne vulgaris	
XX	
PS Example 1; SEQ ID No 23861; 1069pp; English.	
XX	
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic	
CC polypeptides. The proteins and their associated DNA sequences are used in	
CC the treatment, prevention and diagnosis of medical conditions caused by	
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,	
CC pustulosis, hypotosis and osteomyelitis), uveitis and endophthalmitis.	
CC P. acnes is also involved in infections of bone, joints and the central	
CC nervous system, however it is particularly involved in the inflammatory	
CC lesions associated with acne vulgaris. A method for detecting the	
CC presence or absence of P. acnes in a patient comprises contacting a	
CC sample with a binding agent that binds to the proteins of the invention	
CC and determining the amount of bound protein in the sample. The	
CC polypeptides may be used as antigens in the production of antibodies	
CC specific for P. acnes proteins. These antibodies can be used to	
CC downregulate expression and activity of P. acnes polypeptides and	
CC therefore treat P. acnes infections. The antibodies may also be used as	
CC diagnostic agents for determining P. acnes presence, for example, by	
CC enzyme linked immunosorbent assay (ELISA).	
CC Note: The sequence data for this patent did not form part of the printed	
CC specification, but was obtained in electronic format directly from WIPO	
CC at ftp.wipo.int/pub/published_pct_sequences.	
XX	
SQ Sequence 242 AA;	

Alignment Scores:

Pred. No.: 0.0256 Length: 242
 Score: 105.00 Matches: 59
 Percent Similarity: 37.3% Conservative: 15
 Best Local Similarity: 29.80% Mismatches: 68
 Query Match: 9.84% Indels: 57
 DB: 22 Gaps: 11

US-09-920-953-2 (1-598) x AU62666 (1-242)

QY 62 AGAAGCTGTTGATGACCTGGCGCGCAGAGGATGAGCTGGCGGTTGACACCTTCT 121
 DB 1 ArgSerCysProSerArgTrpGluSerProSerAla-----TrpArgProThrProThr 18
 QY 122 AGATAAGTGTGGCTGACCGGAGCTGCTCCCTTCTTCAGTCCCTGGACA----- 175
 DB 19 Profile-----SerSerCysAlaTrpThrProAla 28
 QY 176 -----TGCAAGAGCAGAGATGAGCAGTCAAGTTCATGAGCTTCGTGT 220
 DB 29 ArgGlyAlaAlaSerTrpAspArgLeuArgIleSerSerThrSerArgThrSerAla 48
 QY 221 TTGGCGGAGCAGACCAATAAAG---GCCGAGCATGTACGACGACACGCGCCATCTGG 277
 DB 49 ThrAlaGlyProProArgSerArgProAlaProGlyThrThrSerThrProValThr 68
 QY 278 TCAAGGCGCCGCTGGACCGCCACTTTGACAAGATCAAGAGTACCTGGAGAGA 337
 DB 69 SerArgCys-----ArgSerThrProSerThrProArgSerThrThrProArg 84
 QY 338 CGCTGCAAGAGATGGCGTCAAGCAGGATGTATCCAGACCGCCGCGAGTGGTGGAT 397
 DB 85 AlaSerSerArgProLeuProArgLeuArgThrLysSerProGluProThrSer 104
 QY 398 -----CCACCGCGCAGCAATTTGACTTNCCCA---ACAACTGCGCACCACCACTGTTT 448
 DB 105 HisArgSerGlyAlaSerAsnProThrArgProLeuThrIleCysThrProGluGluThr 124
 QY 449 CATTA-----ACCAACCCCGACCT-----GAGCGC 474
 DB 125 HisProSerGluThrArgProSerProThrArgAlaThrAlaTrpSerLeuThrGluArg 144
 QY 475 TCATTCATCGATTTTGACGGGAGCGGCGCAGTTGC----- 510
 DB 145 ArgHisHisArg-----CysHisArgProAspThrTrpIleSer 157
 QY 511 -----CGAGCGCGCCAGGGGCGCCAGGAGCTGCAAAATCGTTGCCAGC 555
 DB 158 LeuIleAspArgAlaIleLeuGlyPheAlaGlyAlaCysLeuAlaMetAlaSer 175

---SULT 12

ABG05131
 ID ABG05131 standard; Protein; 264 AA.

XX AC ABG05131;
 XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #5122.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX DR N-PSDB; AAS69318.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

XX PS Claim 20; SEQ ID No 35490; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 264 AA;

Alignment Scores:
 Pred. No.: 0.0265 Length: 264
 Score: 105.00 Matches: 59
 Percent Similarity: 35.75% Conservative: 15
 Best Local Similarity: 28.50% Mismatches: 69
 Query Match: 9.84% Indels: 64
 DB: 22 Gaps: 12

US-09-920-953-2 (1-598) x ABG05131 (1-264)

QY 8 CTGACAGCGCGCCACACAGCAGCAGCGGAGTGGCGGCGCAAGAGC 67
 DB 89 LeuGlnAlaGlySerThrAlaPro-----GlyAlaGlyThrProGly 102
 QY 68 TGTGTTGATGACCTGGCGCGCAGAGGATGAGCTGGCGTTTCACACCTTCTACGATA 127
 DB 103 SerArgProThrTrpSerSerSerThrCysSerThrAlaProSerGlyArgAla 122
 QY 128 AGGTGCTGGCTGACCGCGAGCTGCTGCGCTTCTTCAGT----- 166
 DB 123 ArgCys-----AlaCysAlaSerSerSerCysAlaMetSerAlaAlaArg 138
 QY 167 ---CCCTGGACATGCAAGAGC-----AGAGATGAAGCAGG-----TCAGATTCA 208
 DB 139 ArgGlyTrpThrSerProAlaCysTrpArgArgThrSerArgAlaTrpThrThrSer 158
 QY 209 TGAGCTTCGTGTTGGCGGAGCAGACCAATACAAGCGCGGAGCATGTACGACGACACG 268
 DB 159 SerProAlaCysAlaSerSerAlaThrAlaSerValAlaAla----- 172
 QY 269 CCCATCTGCTCAAGGCGCCAGCGCTGGACCGCCACTTTTCACAGATCAAGCAGTACC 328
 DB 173 -----SerThrAlaSerThrTrpProAlaAlaArgThrThr-----GlyGlyThr 187
 QY 329 TTGGAGAGACGCTGCAAGAGATGGCGCTCAAGCAGGATGT-----GATCCAGCA----- 377

Db 188 AlaGlu-SerSerAlaArgProAlaArgAlaSerCysThrGlySerProAlaArgSe 207
 QY 378 -----CGCCGCGGAGTGTGGAGTCCACCGCGAGCAAGATTGACTTNCCEACA 427
 Db 207 rCysTtpArgArgArgProProThrProSerProGlyArg----- 221
 QY 428 ACTGCGCACCAACTGATTTTTCATTAAACCAACCCAGCGCTGAGCGCTCATTCATCGAT 487
 Db 222 -----ProAlaPro-ProSerArgArgThrArgArgA 232
 QY 488 TTTCAGCGGAGCGGAGTGTGGAGTCCACCGCGAGCAAGATTGACTTNCCEACA 547
 Db 232 LaGlnAlaGlyThrSerAlaLeu-----SerProGlyAlaCysPheGlyP 247
 QY 548 TTCCAGCGCTTCTGCTGC 564
 Db 247 roArgSer---CysCys 251

RESULT 13
 BG03533
 J ABG03533 standard; Protein; 900 AA.
 AC ABG03533;
 DT 13-FEB-2002 (first entry)
 XX Novel human diagnostic protein #3524.
 DE Human; Chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 R N-PSDB; AAS67720.

T New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 XX Claim 20; SEQ ID No 33892; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 900 AA;
 SO

Alignment Scores:
 Pred. No.: 0.0474 Length: 900
 Score: 104.50 Matches: 54
 Percent Similarity: 33.16% Conservative: 9
 Best Local Similarity: 28.42% Mismatches: 53
 Query Match: 9.31% Indels: 74
 Gaps: 9
 DB:

US-09-920-953-2 (1-598) x ABG03533 (1-900)

QY 533 CCGTGGCGCCCTGGCGCGCTCGGCA----- 507
 Db 196 ProGlyTrpProProArgLeuArgSerProGlyAsnLeuArgProGlyValGlyGlyLeu 215
 QY 506 ---CTGGCGCTCCCGCT-----CAAAATCGATGGAAT 477
 Db 216 GlyLeuAlaLeuProAlaArgThrAlaAlaAlaProArgProArgGluArgTrpArg 235
 QY 476 GAGCGCTCAGGCTGGGGTTGGTTAATGAAATCAGTTGGTGGCGAGTTGTTGGGNAAG 417
 Db 236 SerPro-----GlyAlaProCysLeuGlyAla 244
 QY 416 TCAATTCGCGCGGTGGAGTCCACCATCCCGCGCGTCTGTGATCTTGTCAAGTGGCGGTGG 357
 Db 245 Gln***ProSerLeuProAlaAlaGluPro-ArgGlyProGlyThrSerGlyArgAl 264
 QY 356 ACGCCCATCTCTTGCAGCGTCTCCAAAGTACTGCTTGTGATCTTGTCAAGTGGCGGTGG 297
 Db 264 aTrpProSerSerAlaAlaArgGlu-----AspCysArgAla----- 277
 QY 296 TCCAGCGGTGGCGCTTGACAGATGGCGGTGCGTGTGATCTTGTGATCTTGTGATCTTGTAT 237
 Db 278 -ProGlyArgGlyProAlaAlaProThrGlyAlaArgArgPro----- 291
 QY 236 TGGTCTGCTCCGCCAACACAGAGCTCATGAACTTGACCTGCTCTTCTCTCTCTTGC 177
 Db 292 -----ProArgProGlyAlaGlyLeuAl 299
 QY 176 ATGTCAGGAGCTCGAAGAGGCA----- 152
 Db 299 aGlyArgProThrArgAlaArgAlaLeuProGlyArgIleGlyArgArgSerProGlyArg 319
 QY 151 -GCAGTCCGGGTGACCGACCTTATCGT-----AGAAGTGTCAACCGCAGC 102
 Db 319 gAlaAlaProHisSerProAlaGlyLeuArgSerAlaAlaArgArgCysProProGly 339
 QY 101 TTCATGCTCTCTGCGC---CGCCAGGT 77
 Db 339 yProCysAlaProArgGlyGlnProGly 348

RESULT 14
 AAU48541
 ID AAU48541 standard; Protein; 305 AA.
 XX
 AC AAU48541;
 XX
 DT 27-FEB-2002 (first entry)
 XX
 DE Propionibacterium acnes immunogenic protein #9437.
 XX
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX

OS Propionibacterium acnes.

PN WO2001181581-A2.

PD 01-NOV-2001.

PP 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

PA SKEIKY YAW, PERSING DH, MITCHAM JL, WANG SS, BHATIA A;

PI L'MAISONNEUVE J, ZHANG Y, JEN S, CARTER D;

XX WPI: 2001-616774/71.

DR N-PSDB; AAS59543.

DR Propionibacterium acnes polypeptides and nucleic acids useful for

vaccinating against and diagnosing infections, especially useful for

treating acne vulgaris

Example 1; SEQ ID NO 9736; 1069pp; English.

Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic

polypeptides. The proteins and their associated DNA sequences are used in

the treatment, prevention and diagnosis of medical conditions caused by

P. acnes. The disorders include SAPHO syndrome (synovitis, acne,

pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.

P. acnes is also involved in infections of bone, joints and the central

nervous system, however it is particularly involved in the inflammatory

lesions associated with acne vulgaris. A method for detecting the

presence or absence of P. acnes in a patient comprises contacting a

sample with a binding agent that binds to the proteins of the invention

and determining the amount of bound protein in the sample. The

polypeptides may be used as antigens in the production of antibodies

specific for P. acnes proteins. These antibodies can be used to

downregulate expression and activity of P. acnes polypeptides and

therefore treat P. acnes infections. The antibodies may also be used as

diagnostic agents for determining P. acnes presence, for example, by

enzyme linked immunosorbent assay (ELISA).

Note: The sequence data for this patent did not form part of the printed

specification, but was obtained in electronic format directly from WIPO

at ftp.wipo.int/pub/published_pct_sequences.

Sequence 305 AA;

Alignment Scores:

Pred. No.: 0.0357 Length: 305

Score: 104.00 Matches: 46

Percent Similarity: 38.67% Conservative: 24

Best Local Similarity: 25.41% Mismatches: 59

Query Match: 9.73% Indels: 52

DB: 22 Gaps: 9

US-09-920-953-2 (1-598) x AAU48541 (1-305)

Db 210 CysLeuAlaSerProProThrSerCys----- 218

QY 212 GCTTCGTGTTTGGCGGAGCAGACCAATACAGGGCGGAAGCATGTACGACGACAGCC 271

Db 219 AlaProCysThrAlaProSer-----SerGlnSerCysThr----- 230

QY 272 ATCTGCTCAAGGGCCACCGCCCTGGACACCGCCACTTTTCACAAAGATCAAGCAGTACCTTGG 331

Db 231 -----AlaThrThrMetThrSerThrThrThrThrProThrGlyAlaSerSer 246

QY 332 GAGAGACGCTGCAAGAGATGGCGGTCACAGCAGGATGTATCCAGCAGCGCCGCGGAGTGG 391

Db 247 ArgThrArgCysSerArgCysGlyGlyLysLysAlaArgProSerProArgProSer--- 265

QY 392 TGGAGTCCACCGCGCAGCAATTTGACTTNCACCAACTGCGCACCCCACTGATTTTCAT 451

Db 266 AlaLeuProProProThrAlaThrThrTyProSerThrProMetThrAlaAsnSerIle 285

QY 452 TAACCCAAACCCGAGCTGAGCGCTCATTCATCCATGATTTTGGAGCGGAGCGCAGTGGC 511

Db 286 CysThrSerProThrLeuSerSer---SerValAspTrpArgProGluGluProIleHis 304

QY 512 GAG 514

Db 305 Glu 305

RESULT 15

ABB65464

ID ABB65464 standard; Protein: 1307 AA.

AC ABB65464;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 23184.

DE Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

OS Drosophila melanogaster.

XX WO200171042-A2.

PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

PI WPI: 2001-656860/75.

DR N-PSDB; ABL09567.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Disclosure; SEQ ID NO 23184; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1307 AA;

Alignment Scores:

Pred. No.: 0.0697 Length: 1307
 Score: 103.50 Matches: 48
 Percent Similarity: 38.95% Conservative: 26
 Best Local Similarity: 25.26% Mismatches: 69
 Query Match: 9.70% Indels: 47
 DB: 22 Gaps: 9

US-09-920-953-2 (1-598) x ABB65464 (1-1307)

QY 56 GGCCAGAGAGCTTTGATGACCTGGCGCGCAGAGCATGAAGCTGGCGTTGACA 115
 Db 451 GlyMetSerThrTyValPrometSerMetAspSerSerGluSerLeuGlyLeuAsp 470
 QY 116 CTTTCTACGATAAG---TGCTGGCTGACCGCGGAGCTGCTCCCTCTTCGAGTCCCTGG 172
 b 471 ProSerThrGluGluValIlePheMetThrAlaAsnArgThrProThrLeuGluProSer 490
 QY 173 ACATGCAAGACAGAGATGAAGCAGGTCAAGTTTCATGAGCTTGGTGTGGCGGAGCAG 232
 Db 491 ValThrSerSerLeuGlyThrSerSerSerAsnProAlaSerAlaIle---GlnPro 509
 QY 233 ACCAATACAGGGCGGAGGATGTACGACGACACGCCCATCTGGTCAAGGGCCACGCC 292
 Db 510 ThrSerSerLysGluHisThrLysThr----- 519
 QY 293 TGGACACCGCCACTTTGACAGATCAAGCAGTACCTTGGAGAGACGCTGCAAGAGATGG 352
 Db 520 ---ThrThrThrThrThrThrThrSerSerSerThr----- 530
 QY 353 GCGTCAAGCAGGATGTGATCCAGCAGCGCCCGGAGTGGTGAGTCCACCGCGACGAAT 412
 Db 531 -----SerThrThrThrThrThrThrProProSerThrGlu 543
 QY 413 TTGACTTCCCACTGCGCAC-----CCAACTGATTTTCAT 451
 Db 544 SerValSerProAsnSerSerAsnAlaAsnThrAsnAlaThrProProAlaProTyHis 563
 QY 452 TAACCCAAACCCCGCTGAGCGCTATTCCATCGATTTGAGCGGGGAGCG----- 502
 Db 564 ---ProArgProGlyIleValLeuAspAspProGluPheLysProGlyGlyArgProArg 582
 QY 503 CCAGTTCCGAGCGCGCCAGGGGGCCAGGAGCGCTCAATCGT-----TTGCCA 553
 Db 583 ProProValGlnArg-----ProProAlaGlnGlnThrLeuProLeuPro 597
 JY 554 GCCCTTGTGATTGAGAGCCATCAGCCA 583
 Db 598 AlaValGlnProThrArgGlnHisLeuPro 607

Search completed: April 16, 2003, 12:58:50
 Job time : 43.5 secs


```

Db      83 ValGlnArgSerLysGlnPheAlaPheLeuAlaTyrAlaLeuGlyGlyAlaSerGluTrp 101
QY      241 AAGGGCGGAAGCATGTACGACGACACGCCCATCTCGTCAAGGCCACGCCCTGGACCAC 300
Db      103 LysGlyLysAspMetArgThrAlaHisLysAspLeuValPro-----HisLeuSerAsp 120
QY      301 CGCCACTTTGACAAAGATCAAGCAGTACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAG 360
Db      121 ValHisPheGlnAlaValAlaAArgHisLeuSerAspThrLeuThrGluLeuGlyValPro 140
QY      361 CAGCATGTGATCCAGACACGCCGCCGGAGGTGGTGGAGTCTCACC CGCCGACGAA 411
Db      141 ProGluAspIleThrAspAlaMetAlaValAlaValAlaSerThrArgThrGlu 157

RESULT 3
A36270
hemoglobin - Tetrahymena pyriformis
C:Species: Tetrahymena pyriformis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: S32555; A36270
R:Takagi, T.; Iwaasa, H.; Yuasa, H.; Shikama, K.; Takemasa, T.; Watanabe,
Biochim. Biophys. Acta 1173, 75-78, 1993
A:Title: Primary structure of Tetrahymena hemoglobins.
A:Reference number: S32555; MUID:93250050; PMID:8485156
A:Accession: S32555
A:Molecule type: mRNA
A:Residues: 1-121 <TAK>
A:Cross-references: EMBL:DL3920; NID:q217409; PIDN:BAA03015.1; PID:q418106
R:Iwaasa, H.; Takagi, T.; Shikama, K.
J. Biol. Chem. 265, 8603-8609, 1990
A:Title: Protozoan hemoglobin from Tetrahymena pyriformis. Isolation, charac
A:Reference number: A36270; MUID:90256780; PMID:3111321
A:Accession: A36270
A:Molecule type: protein
A:Residues: 1-121 <IWA>
C:Genetics:
A:Genetic code: SGC5
C:Superfamily: Tetrahymena globin
C:Keywords: chromoprotein; heme; iron; metalloprotein; monomer
F:1-121/Product: hemoglobin I #status experimental <NARI>
F:3-121/Product: hemoglobin II #status experimental <MAT2>
F:73/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Alignment Scores:
Pred. No.:      2 6e-14      Length:      121
Score:          242.00      Matches:      50
Percent Similarity: 60.87%      Conservative: 20
Best Local Similarity: 43.48%      Mismatches: 41
Query Match:      22.68%      Indels:      4
DB:              1          Gaps:         1

US-09-920-953-2 (1-598) x A36270 (1-121)

QY      67 CTGTTTGATGACCTGGCGCGCGAAGCATGAAGCTGGCGGTGACACCTTCTACGAT 126
Db      7 IleTyrGluLysLeuGlyGlyGluAsnAlaMetLysAlaValProLeuPheTyrLys 26
QY      127 AAGTGCTGCTGACCCGAGCTGCTGCCCTTCTTCGACTCCCTGGACATGCAAGACGAC 186
Db      27 LysValIleAlaAspGluArgValLysHisPhePheLysAsnThrAspMetAspHisGln 46
QY      187 AAGATGAAGCAGGTCAAGTTCATGACGCTCGTGTGGCGGAGCAGACCAATACAGGCG 246
Db      47 ThrLysGlnGlnThrAspPheLeuThrMetLeuLeuGlyGlyProAsnHisTyrLysGly 66
QY      247 CGAAGCATGTACACGACACGCGCCCATCTGTTGTCAGGGCCAGCGGCTGGACACCGCCAC 306
Db      67 LysAsnMetThrGluAlaHis-----LysGlyMetAsnLeuGlnAsnLeuHis 82
QY      307 TTTGACAAGCATCAAGCAGTACCTTTGGAGAGACCGCTGCAAGAGATGGCGTCAAGCAGGAT 366
Db      83 PheAspAlaIleIleGluAsnLeuAlaAlaThrLeuLysGluLeuGlyValThrAspAla 102

```


Best Local Similarity: 34.40% Mismatches: 52
 Query Match: 17.62% Indels: 7
 DB: 2 Gaps: 2

US-09-920-953-2 (1-598) x C70761 (1-136)

QY 67 CTGTTTGTACCTGGCGGCGGCGCAGAGCATGAAGCTGGCGGTGACACCTTCTACGAT 126
 DB 15 IIEYrAspLysIleGlyHisGluAlaIleGluValValGluAspPheTyVal 34
 QY 127 AAGTGTCTGGTACCGGAGTGTCTTCTTCGAGTCCCTGGACATCAAGAGCAG 186
 DB 35 ArgValLeuAlaAspGlnLeuSerAlaPheSerGlyThrAsnMetSerArgLeu 54
 QY 187 AAGTGAAGCAGGTCAAGTTCATGCTTCTGTTTGGCGGAGCAGCAATACAAAGGC 246
 DB 55 LysGlyLysGlnValGluPhePheAlaAlaLeuGlyGlyProGluProTyThrGly 74
 QY 247 CGAAGCATGTACGAGCGACACGCCCATCTGCTCAAGGCCCGGCTGGACCGCCAC 306
 DB 75 AlaPrometLysGlnValHis-----GlnGlyArgGlyIleThrMetHis 90
 QY 307 TTGACAAAGATCAAGCAGTACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGAT 366
 DB 91 PheSerLeuValAlaGlyHisLeuAlaAspAlaLeuThrAlaAlaGlyValProSerGlu 110
 QY 367 GTGATCCAGCAGCGCGGAGTGTGGAGTCCACCGCGAGCAATTGACTTNC----- 421
 DB 111 ThrIleThrGluIleLeuGlyValIle-AlaProLeuAlaValAspValThrSerGlyG1 130
 QY 422 -CCAACAACCTGGC 433
 DB 130 userThrAla 134

RESULT 7
 S60030
 hemoglobin - Paramesidium jenningsi
 C:Species: Paramesidium jenningsi
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: S60030
 R:Yamauchi, K.; Tada, H.; Usuki, I.
 Biochim. Biophys. Acta 1264, 53-62, 1995
 A:Title: Structure and evolution of Paramesidium hemoglobin genes.
 A:Reference number: S60030; MUID:96038820; PMID:7578257
 A:Molecule type: DNA
 A:Residues: 1-117 <YAM>
 .Cross-references: EMBL:D49689; NID:g1071656; PIDN:BAA08540.1; PID:g1384087
 Note: the authors did not translate the codon for residue 1
 ;Genetics:
 A:Gene: Hb
 A:Genetic code: SGC5
 A:Introns: 63/73
 C:Superfamily: Tetrahymena globin
 C:Keywords: chromoprotein; heme; iron; metalloprotein
 F:69/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Alignment Scores:
 Pred. No.: 1.12e-05 Length: 117
 Score: 149.00 Matches: 31
 Percent Similarity: 46.96% Conservativeness: 23
 Best Local Similarity: 26.96% Mismatches: 57
 Query Match: 13.96% Indels: 4
 DB: 1 Gaps: 1

US-09-920-953-2 (1-598) x S60030 (1-117)

QY 67 CTGTTTGTACCTGGCGGCGGCGCAGAGCATGAAGCTGGCGGTGACACCTTCTACGAT 126
 DB 3 LeuPheGluGlnLeuGlyGlyGluAlaAlaValThrAlaValThrGlnPheTyAla 22
 QY 127 AAGTGTCTGCTACCGGAGTGTCTTCTTCGAGTCCCTGGACATGCAAGAGCAG 186
 DB 111 ThrIleThrGluIleLeuGlyValIle-AlaProLeuAlaValAspValThrSerGlyG1 130

DB 23 AsnIleGlnAlaAspAlaThrValAlaAsnPhePheAsnGlyIleAsnMetAlaAspGln 42
 QY 187 AAGTGAAGCAGGTCAAGTTCATGAGCTTCGTTTGGCGGAGCAGCAATACAAAGGC 246
 DB 43 ThrAsnLysThrAlaSerPheLeuGlyGlyProLysAlaTrpGlyGly 62
 QY 247 CGAAGCATGTACGAGCGACACGCCCATCTGCTCAAGGCCCGGCTGGACCGCCAC 306
 DB 63 ArgAsnLeuLysGluValHisAlaAsnMet-----GlyValThrAsnAlaGln 78
 QY 307 TTGACAAAGATCAAGCAGTACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGAT 366
 DB 79 PheThrThrValIleGlyHisLeuArgSerAlaLeuThrSerAlaGlyValAlaAlaAsp 98
 QY 367 GTGATCCAGCAGCGCGGAGTGTGGAGTCCACCGCGAGCA 411
 DB 99 LeuValGluGlnThrValAlaValAlaGluThrValArgGlyAsp 113

RESULT 8

S60032
 hemoglobin - Paramesidium triaurelia
 C:Species: Paramesidium triaurelia
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: S60032
 R:Yamauchi, K.; Tada, H.; Usuki, I.
 Biochim. Biophys. Acta 1264, 53-62, 1995
 A:Title: Structure and evolution of Paramesidium hemoglobin genes.
 A:Reference number: S60030; MUID:96038820; PMID:7578257
 A:Molecule type: DNA
 A:Residues: 1-117 <YAM>
 A:Cross-references: EMBL:D49688; NID:g1071655; PIDN:BAA08539.1; PID:g1384086
 A:Note: the authors did not translate the codon for residue 1
 ;Genetics:
 A:Gene: Hb
 A:Genetic code: SGC5
 A:Introns: 63/73
 C:Superfamily: Tetrahymena globin
 C:Keywords: chromoprotein; heme; iron; metalloprotein
 F:69/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Alignment Scores:
 Pred. No.: 1.12e-05 Length: 117
 Score: 149.00 Matches: 31
 Percent Similarity: 46.96% Conservativeness: 23
 Best Local Similarity: 26.96% Mismatches: 57
 Query Match: 13.96% Indels: 4
 DB: 1 Gaps: 1

US-09-920-953-2 (1-598) x S60032 (1-117)

QY 67 CTGTTTGTACCTGGCGGCGGCGCAGAGCATGAAGCTGGCGGTGACACCTTCTACGAT 126
 DB 3 LeuPheGluGlnLeuGlyGlyGluAlaAlaValThrAlaValThrGlnPheTyAla 22
 QY 127 AAGTGTCTGCTACCGGAGTGTCTTCTTCGAGTCCCTGGACATGCAAGAGCAG 186
 DB 23 AsnIleGlnAlaAspAlaThrValAlaAsnPhePheAsnGlyIleAsnMetAlaAspGln 42
 QY 187 AAGTGAAGCAGGTCAAGTTCATGAGCTTCGTTTGGCGGAGCAGCAATACAAAGGC 246
 DB 43 ThrAsnLysThrAlaSerPheLeuGlyGlyProLysAlaTrpGlyGly 62
 QY 247 CGAAGCATGTACGAGCGACACGCCCATCTGCTCAAGGCCCGGCTGGACCGCCAC 306
 DB 63 ArgAsnLeuLysGluValHisAlaAsnMet-----GlyValThrAsnAlaGln 78
 QY 307 TTGACAAAGATCAAGCAGTACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGAT 366
 DB 79 PheThrThrValIleGlyHisLeuArgSerAlaLeuThrSerAlaGlyValAlaAlaAsp 98
 QY 367 GTGATCCAGCAGCGCGGAGTGTGGAGTCCACCGCGAGCA 411

Db 99 LeuValGluGlnThrValAlaValAlaGluThrValArgGlyasp 113

RESULT 9
S27185
hemoglobin major component - Paramesicium caudatum
N;Alternate names: B-type hemoglobin; monomeric hemoglobin; myoglobin
C;Species: Paramesicium caudatum
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: S27185; S05230; JQ2009; JQ1316
R;Yamauchi, K.; Ochiai, T.; Usuki, I.
Biochim. Biophys. Acta 1171, 81-87, 1992
A;Title: The unique structure of the Paramesicium caudatum hemoglobin gene: the presence of
A;Reference number: S27185; MUID:93042011; PMID:1420365
A;Molecule type: DNA
A;Accession: S27185
A;Residues: 1-116 <VAM>
A;Cross-references: EMBL:M99047
A;Note: the authors did not translate the codon for residue 1
R;Iwasawa, H.; Takagi, T.; Shikama, K.
J. Mol. Biol. 208, 355-358, 1989
;Title: Protozoan myoglobin from Paramesicium caudatum. Its unusual amino acid sequence.
;Reference number: S05230; MUID:89362481; PMID:2769763
A;Accession: S05230
A;Molecule type: protein
A;Residues: 2-117 <IWA>
A;Experimental source: strain syngen 3 stock StG1
R;Yamauchi, K.; Tada, H.; Ochiai, T.; Usuki, I.
Gene 126, 243-246, 1993
A;Title: Structure of the Paramesicium caudatum gene encoding the B-type of the major hemo
A;Reference number: JQ2009; MUID:93246250; PMID:8482540
A;Accession: JQ2009
A;Molecule type: DNA
A;Residues: 1-117 <VAM>
A;Cross-references: DDBJ:D12916; NID:g397832; PIDN:BAA02300.1; PID:g415308
A;Experimental source: stock K33
R;Yamauchi, K.; Mukai, M.; Ochiai, T.; Usuki, I.
Biochim. Biophys. Res. Commun. 182, 195-200, 1992
A;Title: Molecular cloning of the cDNA for the major hemoglobin component from Paramesi
A;Reference number: JQ1316; MUID:92118012; PMID:1731779
A;Accession: JQ1316
A;Molecule type: mRNA
A;Residues: 1-117 <YNA>
A;Cross-references: GB:M57542
C;Genetics:
A;Gene: Hb
A;Genetic code: SGC5
A;Introns: 63/3
C;Superfamily: Tetrahymena globin
C;Keywords: acetylated amino end; chromoprotein; heme; iron; metalloprotein; oxygen carr
;2/117/Product: hemoglobin major component #status experimental <MAT>
;2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
F;69/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Alignment Scores:
Pred. No.: 1-72e-05 Length: 117
Score: 147.00 Matches: 31
Percent Similarity: 46.96% Conservative: 23
Best Local Similarity: 26.96% Mismatches: 57
Query Match: 13.78% Indels: 4
DB: Gaps: 1

US-09-920-953-2 (1-598) x S27185 (1-117)

Oy 67 CTGTTTGTACCTGGCGCGAGAGCATGAAGCTGGCGGTGGACACTTCATCCAT 126
||||| : ||||| :
3 LeupheGluGlnLeuGlyGlnAlaAlaValThraAlaGlnPheTyraLa 22

Oy 127 AAGTGCTGGCTGACC CGGAGCTCTGCCCTTCCTGAGTCCCTGGACATCAAGACGAG 186
||| : ||||| :
23 AsnleGlnAlaAspAlaThrValAlaThrPhePheAsnGlylleAspMetProAsnGln 42

Oy 187 AAGATGAAGCAGGTCACAGTTTCATGAGCTTCGTTGGCGGAGCACACCAATACAAGGC 246
||| : ||||| :
43 ThrAsplystThraAlaAlaPheLeuCysAlaAlaLeuGlyGlyProAsnAlaTrpAlaGly 62

Oy 247 CGAAGCATGTACGACGACGACGCCCCATCTGTCGTAAGGGCCCGCTGGACCCGCCAC 306
||| : ||||| :
63 ArgAsnLeuLysGluValHisaAlaAsnMet-----GlyValThrAsnAlaGln 78

US-09-920-953-2 (1-598) x S27185 (1-117)

Oy 67 CTGTTTGTACCTGGCGCGAGAGCATGAAGCTGGCGGTGGACACTTCATCCAT 126
||||| : ||||| :
3 LeupheGluGlnLeuGlyGlnAlaAlaValThraAlaGlnPheTyraLa 22

Oy 127 AAGTGCTGGCTGACC CGGAGCTCTGCCCTTCCTGAGTCCCTGGACATCAAGACGAG 186
||| : ||||| :
23 AsnleGlnAlaAspAlaThrValAlaThrPhePheAsnGlylleAspMetProAsnGln 42

Oy 187 AAGATGAAGCAGGTCACAGTTTCATGAGCTTCGTTGGCGGAGCACACCAATACAAGGC 246
||| : ||||| :
43 ThrAsplystThraAlaAlaPheLeuCysAlaAlaLeuGlyGlyProAsnAlaTrpAlaGly 62

Oy 247 CGAAGCATGTACGACGACGACGCCCCATCTGTCGTAAGGGCCCGCTGGACCCGCCAC 306
||| : ||||| :
63 ArgAsnLeuLysGluValHisaAlaAsnMet-----GlyValThrAsnAlaGln 78

US-09-920-953-2 (1-598) x S60031 (1-118)

Oy 67 CTGTTTGTACCTGGCGCGAGAGCATGAAGCTGGCGGTGGACACTTCATCCAT 126
||||| : ||||| :
3 LeupheGluGlnLeuGlyGlnAlaAlaValThraAlaGlnPheTyraLa 22

Oy 127 AAGTGCTGGCTGACC CGGAGCTCTGCCCTTCCTGAGTCCCTGGACATCAAGACGAG 186
||| : ||||| :
23 AsnleGlnAlaAspAlaThrValAlaThrPhePheAsnGlylleAsnGlyProThrGln 42

Oy 187 AAGATGAAGCAGGTCACAGTTTCATGAGCTTCGTTGGCGGAGCACACCAATACAAGGC 246
||| : ||||| :
43 ThrAsplystThraAlaAlaPheLeuCysAlaAlaLeuGlyGlyProAsnAlaTrpAlaGly 62

Oy 247 CGAAGCATGTACGACGACGACGCCCCATCTGTCGTAAGGGCCCGCTGGACCCGCCAC 306
||| : ||||| :
63 ArgAsnLeuLysGluValHisaAlaAsnMet-----GlyValThrAsnAlaGln 78

US-09-920-953-2 (1-598) x S60031 (1-118)

Oy 67 CTGTTTGTACCTGGCGCGAGAGCATGAAGCTGGCGGTGGACACTTCATCCAT 126
||||| : ||||| :
3 LeupheGluGlnLeuGlyGlnAlaAlaValThraAlaGlnPheTyraLa 22

Oy 127 AAGTGCTGGCTGACC CGGAGCTCTGCCCTTCCTGAGTCCCTGGACATCAAGACGAG 186
||| : ||||| :
23 AsnleGlnAlaAspAlaThrValAlaThrPhePheAsnGlylleAsnGlyProThrGln 42

Oy 187 AAGATGAAGCAGGTCACAGTTTCATGAGCTTCGTTGGCGGAGCACACCAATACAAGGC 246
||| : ||||| :
43 ThrAsplystThraAlaAlaPheLeuCysAlaAlaLeuGlyGlyProAsnAlaTrpAlaGly 62

Oy 247 CGAAGCATGTACGACGACGACGCCCCATCTGTCGTAAGGGCCCGCTGGACCCGCCAC 306
||| : ||||| :
63 ArgAsnLeuLysGluValHisaAlaAsnMet-----GlyValThrAsnAlaGln 78

US-09-920-953-2 (1-598) x S60031 (1-118)

Oy 67 CTGTTTGTACCTGGCGCGAGAGCATGAAGCTGGCGGTGGACACTTCATCCAT 126
||||| : ||||| :
3 LeupheGluGlnLeuGlyGlnAlaAlaValThraAlaGlnPheTyraLa 22

Oy 127 AAGTGCTGGCTGACC CGGAGCTCTGCCCTTCCTGAGTCCCTGGACATCAAGACGAG 186
||| : ||||| :
23 AsnleGlnAlaAspAlaThrValAlaThrPhePheAsnGlylleAsnGlyProThrGln 42

Oy 187 AAGATGAAGCAGGTCACAGTTTCATGAGCTTCGTTGGCGGAGCACACCAATACAAGGC 246
||| : ||||| :
43 ThrAsplystThraAlaAlaPheLeuCysAlaAlaLeuGlyGlyProAsnAlaTrpAlaGly 62

Oy 247 CGAAGCATGTACGACGACGACGCCCCATCTGTCGTAAGGGCCCGCTGGACCCGCCAC 306
||| : ||||| :
63 ArgAsnLeuLysGluValHisaAlaAsnMet-----GlyValThrAsnAlaGln 78

US-09-920-953-2 (1-598) x S60031 (1-118)

Oy 67 CTGTTTGTACCTGGCGCGAGAGCATGAAGCTGGCGGTGGACACTTCATCCAT 126
||||| : ||||| :
3 LeupheGluGlnLeuGlyGlnAlaAlaValThraAlaGlnPheTyraLa 22

Oy 127 AAGTGCTGGCTGACC CGGAGCTCTGCCCTTCCTGAGTCCCTGGACATCAAGACGAG 186
||| : ||||| :
23 AsnleGlnAlaAspAlaThrValAlaThrPhePheAsnGlylleAsnGlyProThrGln 42

Oy 187 AAGATGAAGCAGGTCACAGTTTCATGAGCTTCGTTGGCGGAGCACACCAATACAAGGC 246
||| : ||||| :
43 ThrAsplystThraAlaAlaPheLeuCysAlaAlaLeuGlyGlyProAsnAlaTrpAlaGly 62

Oy 247 CGAAGCATGTACGACGACGACGCCCCATCTGTCGTAAGGGCCCGCTGGACCCGCCAC 306
||| : ||||| :
63 ArgAsnLeuLysGluValHisaAlaAsnMet-----GlyValThrAsnAlaGln 78

US-09-920-953-2 (1-598) x S60031 (1-118)

Oy 67 CTGTTTGTACCTGGCGCGAGAGCATGAAGCTGGCGGTGGACACTTCATCCAT 126
||||| : ||||| :
3 LeupheGluGlnLeuGlyGlnAlaAlaValThraAlaGlnPheTyraLa 22

Oy 127 AAGTGCTGGCTGACC CGGAGCTCTGCCCTTCCTGAGTCCCTGGACATCAAGACGAG 186
||| : ||||| :
23 AsnleGlnAlaAspAlaThrValAlaThrPhePheAsnGlylleAsnGlyProThrGln 42

Oy 187 AAGATGAAGCAGGTCACAGTTTCATGAGCTTCGTTGGCGGAGCACACCAATACAAGGC 246
||| : ||||| :
43 ThrAsplystThraAlaAlaPheLeuCysAlaAlaLeuGlyGlyProAsnAlaTrpAlaGly 62

Oy 247 CGAAGCATGTACGACGACGACGCCCCATCTGTCGTAAGGGCCCGCTGGACCCGCCAC 306
||| : ||||| :
63 ArgAsnLeuLysGluValHisaAlaAsnMet-----GlyValThrAsnAlaGln 78

US-09-920-953-2 (1-598) x S60031 (1-118)

Oy 67 CTGTTTGTACCTGGCGCGAGAGCATGAAGCTGGCGGTGGACACTTCATCCAT 126

[illegible]

Qy 509 -----GCCGAGCGCCCGCAGGGGCCAGGCCTTCCA-----AATCGT 547
||||| ||||| ||||| |||||
Db 1067 LeuCysAlaGluProAlaGluGlyAlaProArgProAlaTrpAspProArgArgGlyGly 1086

Qy 548 TTGGCCACCCCTTGCTGCATTGAAGAGCACATCAG-----CCATTTTCGCAC 592
||| ||||| ||| :|||::: ||| |||||
Db 1087 LeuSerAlaLeuLeuAlaAlaPheAlaHisArgLeuLeuGlyCysThrProAspSerHis 1104

RESULT 13
S32975
gene BCRF2 protein - human herpesvirus 4
N:Alternate names: gene BWRF1 protein
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Nov-2000
C:Accession: S32975; S02383; S32976; S32977; S32978; S32979; S32980; S32981; S32982; S32983
R:Farrell, P.J.
submitted to the EMBL Data Library, March 1988
A:Reference number: S32973
A:Accession: S32975
A:Molecule type: DNA
:Residues: 1-383 <PAR>
J: PID:g1334846; PID:g1334847
A:Note: each of the two authors supplied translations in EMBL:V01555 for this repeated
R:Walls, D.; Cannon, F.
EMBO J. 7, 1191-1196, 1988
A:Title: The expression of novel antigens from the Epstein-Barr virus large internal repeat
A:Reference number: S02381; MUID:88296424; PMID:2841116
A:Accession: S02383
A:Molecule type: DNA
A:Residues: 88-144 <WAL>
A:CROSS-references: EMBL:X07816
C:Genetics:
A:Gene: BCRF2_1; BWRF1_2; BWRF1_3; BWRF1_4; BWRF1_5; BWRF1_6; BWRF1_7; BWRF1_8; BWRF1_9;
A:Note: twelve consecutive ORFs apparently encode the identical polypeptide
C:Superfamily: proline-rich protein
C:Keywords: membrane protein; surface antigen

Alignment Scores:
Pred No.: 0.172 Length: 383
Score: 104.50 Matches: 67
Percent Similarity: 33.76% Conservative: 13
Best Local Similarity: 28.27% Mismatches: 76
Query Match: 9.79% Indels: 82
DB: 2 Gaps: 11

US-09-920-953-2 (1-598) x S32975 (1-383)
Y 20 CCACCAGCACCACAGACGCGCGAAGCGGATCGGGCCGCAAGACTGTTCATGACC 79
||| ||||| ||| :|||
D 62 ProGlnAlaProGlnThrArg-----ValSerAla 71

Qy 80 TGGCGGGCGCAGAGCATGAAAGCTGGCGGTGGACACTTCTACGATAAGGTGCTGCTG 139
||| :||| ||||| ||||| :||| |||
Db 72 SerArgAlaAspArgProArgAlaTrpArgLeuLeuGlyAlaSerArgGlyTrpPhe 91
||| ||||| ||||| ||||| |||||
Qy 140 ACCCGAGCTGCTGCCCTTCCTCGAGTCCCTGGA----- 173
||| ||||| ||||| ||||| |||||
Db 92 CysProSerLeuCysProSerGluGlu-ProGlyThrSerGlyThrProGluProLeuGl 111
||| ||||| ||||| ||||| |||||
Qy 173 ----- 173
Db 111 yProAlaSerArgArgProProGlyLeuArgSerProLeuSerProValLysProLysgl 131
||| ||||| ||||| ||||| |||||
Qy 174 -----CATGCAGACGAGACATGAAAGCGTCAAGTTCATGA 211
||| ||||| ||||| ||||| |||||
Db 131 uCysLeuArgGlyAlaThrLeuGlyAlaGlnAlaProGluSerArgGlyGlnGlyHis-- 150
||| ||||| ||||| ||||| |||||
Qy 212 GCTTCGTGTTGGCGGACGACGACAATACAA-----GGCGGAAGCATGTACGACGCAC 265
||| ||||| ||||| ||||| |||||
Db 151 -LeuArgValProProArgValProGlyGlnProGluGlyProArg-----GlnProGl 168
||| ||||| ||||| ||||| |||||
Qy 266 ACGCCCATCTGGTCAAGGCCGACGCGCTGGACCA----- 299

Search completed: April 16, 2003, 13:01:58
Job time : 29 secs


```
OY 361 CAGGATGTGATCCAGCAGCGCGGAGTGGTGGAGTCCACCGCAGGAA 411
Db 141 ProGluAspIleThrAspAlaMetAlaValAlaSerThrArgThrGlu 157

RESULT 3
GLB_TETPY
ID GLB_TETPY STANDARD; PRT; 121 AA.
AC P17724;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myoglobin (Hemoglobin).
OS Tetrahymena pyriformis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5908;
RN [1]
RN MEDLINE=90256780; PubMed=2111321;
RX Iwasa H., Takagi T., Shikama K.;
RA Takagi T., Iwasa H., Yuasa H., Shikama K., Takemasa T.,
YA Watanabe Y.;
RT "Primary structure of Tetrahymena hemoglobins.";
RL Biochim. Biophys. Acta 1173:75-78(1993).
RN [2]
RN MEDLINE=90256780; PubMed=2111321;
RX Iwasa H., Takagi T., Shikama K.;
RA Takagi T., Iwasa H., Yuasa H., Shikama K., Takemasa T.,
YA Watanabe Y.;
RT "Primary structure of Tetrahymena hemoglobins.";
RL Biochim. Biophys. Acta 1173:75-78(1993).
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN
FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D13920; BAA03015.1;
DR PIR: A36270; A36270.
DR PIR: S32555; S32555.
DR InterPro: IPR001486; Bac_globin.
DR Pfam: PF01152; Bac_globin; 1.
DR ProDom: PD004840; Bac_globin; 1.
R PROSITE: PS01213; GLOBIN_FAM_2; 1.
AW Heme; Oxygen transport; Transprot; Acetylation.
FT MOD_RES 1 1 ACETYLATION (PROBABLE).
FT METAL 73 73 IRON (HEME PROXIMAL LIGAND) (POTENTIAL).
SQ SEQUENCE 121 AA; 13684 MW; 6D02948060CA2A5E CRC64;

Alignment Scores:
Pred. No.: 3.43e-13 Length: 121
Score: 242.00 Matches: 50
Percent Similarity: 60.87% Conservative: 20
Best Local Similarity: 43.48% Mismatches: 41
Query Match: 22.68% Indels: 4
DB: 1 Gaps: 1

US-09-920-953-2 (1-598) x GLB_TETPY (1-121)

OY 67 CTGTTTGTATCCCTGGCGGCAGAGGATGAAGTGGGTTGACACCTTCTACGAT 126
Db 7 IleTyrGluLysLeuGlyGlyGluAsnAlaMetLysAlaAlaValProLeuPheTyrLys 26

OY 127 AAGTGTGCTGCTACCGGAGCTGCTGCCCTTCTTCGAGTCCCTGGACATCAAGACGAC 186
Db 27 LysValLeuAlaAspGluArgValLysHisPhePheLysAsnThrAspMetAspHisGln 46
```

```
OY 187 AAGATGAAGCAGGCTCAAGTTCATGCTTCGTTTGGCGGAGCAGACCAATACAAGGC 246
Db 47 ThrLysGlnGlnThrAspPheLeuThrMetLeuLeuGlyGlyProAsnHisTyrLysGly 66

OY 247 CGAAGCATGTACGACGACGCGCCATCTGTGTCGAAGGCGCCAGCGCTGGACCCGAC 306
Db 67 LysAsnMetThrGluAlaHis-----LysGlyMetAsnLeuGlnAsnLeuHis 82

OY 307 TTTTCACAAGATCAAGCAGTACCTTGGAGAGAGCGTGCACAGAGATGGCGGTCAAGCAGAT 366
Db 83 PheAspAlaIleIleGluAsnLeuAlaAlaThrLeuLysGluLeuGlyValThrAspAla 102

OY 367 GTGATCCAGCAGCGCGGAGTGGTGGAGTCCACCGCGCAGGAA 411
Db 103 ValIleAsnGluAlaAlaLysValIleGluHisThrArgLysAsp 117

RESULT 4
GLB2_CHLEU
ID GLB2_CHLEU STANDARD; PRT; 171 AA.
AC P52334;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Globin Li410.
GN Li410.
OS Chlamydomonas eugametos.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3053;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=UTEX 9;
RX MEDLINE=94232186; PubMed=8177215;
RA Couture M., Chamberland H., St Pierre B., Lafontaine J., Guertin M.;
RT "Nuclear genes encoding chloroplast hemoglobins in the unicellular
RL green alga Chlamydomonas eugametos.";
CC Mol. Gen. Genet. 243:185-197(1994).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST, PARTICULARLY IN THE PYRENOID
CC AND THE THYLAKOID REGION.
CC -1- INDUCTION: BY LIGHT.
CC -1- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X72915; CAA51420.1;
DR InterPro: IPR001486; Bac_globin.
DR Pfam: PF01152; Bac_globin; 1.
DR ProDom: PD004840; Bac_globin; 1.
DR PROSITE: PS01213; GLOBIN_FAM_2; 1.
KW Heme; Oxygen transport; Transprot; Chloroplast.
FT METAL 63 63 IRON (HEME DISTAL LIGAND) (BY
SIMILARITY).
FT METAL 111 111 IRON (HEME PROXIMAL LIGAND) (POTENTIAL).
SQ SEQUENCE 171 AA; 18598 MW; 236E759D0C2D42FA CRC64;

Alignment Scores:
Pred. No.: 3.9e-13 Length: 171
Score: 241.50 Matches: 56
Percent Similarity: 57.58% Conservative: 20
Best Local Similarity: 42.42% Mismatches: 53
Query Match: 22.63% Indels: 3
DB: 1 Gaps: 2

US-09-920-953-2 (1-598) x GLB2_CHLEU (1-171)

OY 16 GCGGCCACCACGACGAGCGGCGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGTTGAT 75
```

```

|||||
29 AlaAlaThrAlaThrThrProAlaProThrLysLysCys---ProPheSerLeuPheAla 47
QY 76 GACCTGGCGGCGAGAGCATGAAGCTGGGGTGTACACCTTCTACGATAAGTGTGTG 135
Db 48 LysLeuGlyGlyArgGluAlaValGluAlaAlaValAspLysPheTyrAsnLysValVal 67
QY 136 GCTGACCGGAGCTGCTGCCCTTCTTCAGTCCCTGAGTCCATCAAGACAGCAAGATGAAG 195
Db 68 AlaAspProThrValSerValPhePheSerLysThrAspMetLysValGlnArgSerLys 87
QY 196 CAGGTCAAGTTCATGAGTCTGCTGTTGGCGGAGCAGACCAATCAAGAGCGGCGGACCATG 255
Db 88 GlnPheAlaPheLeuAlaValAlaLeuGlyGlyAlaAlaGluTrpLysGlyLysAspMet 107
QY 256 TACGACGACGACGCCCTGCTGTCAGGCGCCAGCGCTGGACCGCCACTTTCACAAAG 315
Db 108 ArgThrAlaHisLysAspLeuValPro-----HisLeuThrAspValHisPheGlnAla 125
" 316 ATCAAGCAGTACCTTGGAGACAGCTGCAAGAGATGGCGGTCAAGCAGGATGTGTCCAG 375
Db 126 ValValArgHisLeuSerAspThrLeuAlaGluLeuGlyValThrProGlyAspIleAla 145
QY 376 CACGCGCGCGGAGTGTGAGTCCACCGCGGACGAA 411
Db 146 AspAlaMetAlaValAlaSerThrLysThrGlu 157

```

RESULT 5

```

GLB_TETTH STANDARD; PRT; 121 AA.
AC Q03459;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myoglobin (Hemoglobin).
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93250050; PubMed=8485156;
RA Takagi T., Iwasa H., Yuasa H., Shikama K., Takemasa T.,
RA Watanabe Y.;
RA "Primary structure of Tetrahymena hemoglobins.";
RL Biochim. Biophys. Acta 1173:75-78(1993).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN
FAMILY.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC EMBL: D13919; BAA03014.1;
CC PIR: S32556; S32556.
DR InterPro: IPR001486; Bac_globin.
DR Pfam: PF01152; Bac_globin; 1.
DR PRODOM: PD004840; Bac_globin; 1.
DR PROSITE: PS01213; GLOBIN.FAM.2; 1.
KW Heme; Oxygen transport; Transport; Acetylation.
FT MOD_RES 1 73 1 ACETYLATION (PROBABLE).
FT METAL 73 73 IRON (HEME PROXIMAL LIGAND) (POTENTIAL).
SO SEQUENCE 121 AA; 13590 MW; 5FC4743192605EDF CRC64;

```

Alignment Scores:

```

Pred No.: 3.66e-12 Length: 121
Score: 230.00 Matches: 47
Percent Similarity: 60.87% Conservative: 23

```

```

Best Local Similarity: 40.87% Mismatches: 41
Query Match: 21.56% Indels: 4
DB: 1 Gaps: 1
US-09-920-953-2 (1-598) x GLB_TETTH (1-121)
QY 67 CTGTTGTGATGACCTGGCGGCGGCGAGCATGAAGCTGGGGTGTACACCTTCTACCAT 126
Db 7 ValPheGluLysLeuGlyGlyGlnAlaAlaMetHisAlaAlaValProLeuPheTyrLys 26
QY 127 AAGTGTGCTGCTGACCGGAGCTGCTGCCCTTCTTCAGTCCCTGAGTCCATCAAGAGCAG 186
Db 27 LysValLeuAlaAspAspArgValLysHisTyrPheLysAsnThrAsnMetGluHisGln 46
QY 187 AAGTGAAGCAGGTCAAGTTCATGAGTCTGCTGTTGGCGGAGCAGACCAATACAGGCG 246
Db 47 AlalysGlnGlnGluAspPheLeuThrMetLeuGlyGlyProAsnHisTyrLysGly 66
QY 247 CGAAGCATGTACGACGACGACGCCCATCTGCTCAAGGCGCCAGCGCTGGACCCGCCAC 306
Db 67 LysAsnMetAlaGluAlaHis-----LysGlyMetAsnLeuGlnAsnSerHis 82
QY 307 TTGACAAGATCAAGCAGTACTCTTGGAGAGAGCTGCAAGAGATGGGCTCAAGCAGGAT 366
Db 83 PheAspAlaIleIleGluAsnLeuAlaAlaThrLeuLysGluLeuGlyValSerAspGln 102
QY 367 GTGATCCAGCAGCGCGGAGTGTGAGTCCACCGCGGACGAA 411
Db 103 IleIleGlyGluAlaAlaLysValIleGluHisThrArgLysAsp 117

```

RESULT 6

```

GLBN_MYCTU STANDARD; PRT; 136 AA.
AC Q10784;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hemoglobin-like protein HbN (Flavohemoglobin).
GN GLBN OR RV1542C OR MT1594 OR MTCV48.23.
OS Mycobacterium tuberculosis,
OS Mycobacterium bovis, and
OS Mycobacterium smegmatis.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773; 1765, 1772;
RN [1]
RP SEQUENCE FROM N.A.
RX SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

```

RC SPECIES=M.bovis; STRAIN=BCG;
RX MEDLINE=99432214; PubMed=10500158;
RA Couture M., Yeh S.R., Wittenberg B.A., Wittenberg J.B., Ouellet Y.,
RA Rousseau D.L., Guertin M.;
RT "A cooperative oxygen-binding hemoglobin from Mycobacterium
RT tuberculosis";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11223-11228(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=M.smegmatis; STRAIN=LR222;
RA Kumar A., Rawat V.P.S., Das R.H.;
RA Yeh S.R., Couture M., Ouellet Y., Rousseau D.L.;
RT "Cloning, expression and characterization of globin like genes of
RT Mycobacterium in insect cells(sf9).";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=20102674; PubMed=10636862;
RA Yeh S.R., Couture M., Ouellet Y., Rousseau D.L.;
RT "A cooperative oxygen binding hemoglobin from Mycobacterium
RT tuberculosis. Stabilization of heme ligands by a distal tyrosine
T residue.";
J. Biol. Chem. 275:1679-1684(2000).
CC -I- FUNCTION: Binds oxygen cooperatively with very high affinity
CC (P50) = 0.013 mmHg at 20 degrees Celsius) because of a fast
CC combination (25 microm(-1).s(-1)) and a slow dissociation (0.2 s(-
CC 1)) rate.
CC -I- SUBUNIT: Homodimer.
CC -I- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 274020; CAA98320.1; .
DR EMBL: AE007025; AAK45860.1; .
DR EMBL: AF130980; AAD28758.1; .
DR EMBL: AJ249386; CAB56291.1; .
DR TIGR: Wt1594; .
DR TubercuList; Rv1542c; .
DR InterPro: IPR001486; Bac_globin.
DR Pfam: PF01152; Bac_globin; 1.
DR ProDom: PD004840; Bac_globin; 1.
DR PROSITE: PS01213; GLOBIN_FAM_2; 1.
W Heme; Oxygen transport; Transport; Complete proteome.
F METAL 81 81 IRON (HEME PROXIMAL LIGAND) (POTENTIAL).
SQ SEQUENCE 136 AA; 1448 MW; B75D01A45BC0645B CRC64;
Alignment Scores:
Pred. No.: 1,48e-08 Length: 136
Score: 188.00 Matches: 43
Percent Similarity: 52.80% Conservatives: 23
Best Local Similarity: 34.40% Mismatches: 52
Query Match: Indels: 7
DB: Gaps: 2

US-09-920-953-2 (1-598) x GLBN_MYCTU (1-136)

QY 67 CTGTTTGTACCTGGCGGCGGAGAGGATGAGTGGGTTGACCTTCTACGAT 126
Db 15 IletyrAspLysIleGlyHisGluAlaIleGluValValGluAspPheTyrVal 34
QY 127 AAGTGTCTGCTGACCCGAGCTGCTGCTTCTTTCAGTCCCTGGACATCCAGACGAG 186
Db 35 ArgValLeuAlaAspAspGlnLeuSerAlaPhePheSerGlyThrAsnMetSerArgLeu 54
QY 187 AAGATGAGCAGGTCAAGTTTCATGAGCTTCGTGTTTGGCGGAGCAGACCAATACAGGCG 246
Db 55 LysGlyLysGlnValGluPhePheAlaAlaLeuGlyGlyProGluProTyrThrGly 74

QY 247 CGAAGCATGTACGACGACACGCCATCTGTGTCAGGCGGCGGCTGGACACGCCAC 306
Db 75 AlaProMetLysGlnValHis-----GlnGlyArgGlyIleThrMetHisHis 90
QY 307 TTTGACAAGATCAACAGCTACCTTGGAGAGAGCTGCAAGAGATGGCGCTCAAGCAGAT 366
Db 91 PheSerLeuValAlaGlyHisLeuAlaAspAlaLeuThrAlaAlaGlyValProSerGlu 110
QY 367 GTGATCCAGCAGCGCGGAGTGTGGAGTCCACCGGACGAAATTTGACTTNC----- 421
Db 111 ThrIleThrGluIleLeuGlyValIle-AlaProLeuAlaValAspValThrSerGlyGlu 130
QY 422 -CCAACAACCTGCG 433
Db 130 uSerThrThrAla 134
RESULT 7
GLBN_NOSSN STANDARD; PRT; 118 AA.
AC P52335;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cyanogloblin.
GN GLBN
OS Nostoc sp. (strain MUN 8820).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=55397;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97086627; PubMed=8932316;
RA Hill D.R., Belbin T.J., Thorsteinsson M.V., Bassam D., Brass S.,
RA Ernst A., Boger P., Paerl H., Mulligan M.E., Potts M.;
RT "GLN (cyanogloblin) is a peripheral membrane protein that is
RT restricted to certain Nostoc spp.";
RL J. Bacteriol. 178:6587-6598(1996).
CC -I- SUBUNIT: MONOMER.
CC -I- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L47979; AAB41122.1; .
DR InterPro: IPR001486; Bac_globin.
DR Pfam: PF01152; Bac_globin; 1.
DR ProDom: PD004840; Bac_globin; 1.
DR PROSITE: PS01213; GLOBIN_FAM_2; 1.
KW Heme; Oxygen transport; Transport.
FT METAL 70 70 IRON (HEME PROXIMAL LIGAND) (POTENTIAL).
SQ SEQUENCE 118 AA; 12794 MW; D5B3E461FE4C860C CRC64;
Alignment Scores:
Pred. No.: 1.37e-06 Length: 118
Score: 165.00 Matches: 39
Percent Similarity: 45.45% Conservatives: 16
Best Local Similarity: 32.23% Mismatches: 52
Query Match: Indels: 14
DB: Gaps: 2

US-09-920-953-2 (1-598) x GLBN_NOSSN (1-118)

QY 67 CTGTTTGTACCTGGCGGCGGAGAGGATGAGTGGGTTGACCTTCTACGAT 126
Db 4 LeuTyrAspAsnIleGlyGlyGlnProAlaIleGluValValAspGluLeuHisLys 23

RA Yamauchi K., Tada H., Ochiai T., Usuki I.;
 RT "Structure of the Paramesium caudatum gene encoding the B-type of the
 RL major hemoglobin component.";
 RL Gene 126:243-246(1993).
 RN [4]
 RP SEQUENCE.
 RX MEDLINE=89362481; PubMed=2769763;
 RA Iwasa H., Takagi T., Shikama K.;
 RT "Protozoan myoglobin from Paramesium caudatum. Its unusual amino acid
 sequence.";
 RL J. Mol. Biol. 208:355-358(1989).
 CC -!- SUBUNIT: MONOMER.
 CC -!- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN
 CC FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; S49254; AAB24268.2;
 DR EMBL; M99047; AAA29447.2;
 DR EMBL; M57542; AAA29446.2;
 DR EMBL; D12916; BAA02300.1;
 DR PIR; S05230; S05230.
 DR PIR; S27185; S27185.
 DR PIR; JQ2009; JQ2009.
 DR InterPro; IPR001486; Bac_globin.
 DR Pfam; PF01152; Bac_globin; 1.
 DR ProDom; PD004840; Bac_globin; 1.
 DR PROSITE; PS01213; GLOBIN_FAM_2; 1.
 KW Heme: Oxygen transport; Transport; Acetylation.
 FT INIT MET 0
 FT MOD_RES 1
 FT METAL 68
 FT CONFLICT 17
 FT ACETYLATION.
 FT IRON (HEME PROXIMAL LIGAND) (POTENTIAL).
 FT A -> D (IN REF.2).
 SQ SEQUENCE 116 AA; 11907 MW; AB46AE315F44855 CRC64;
 Alignment Scores:
 Pred. No.: 4,78e-05 Length: 116
 Score: 147.00 Matches: 31
 Percent Similarity: 46.96% Conservative: 23
 Best Local Similarity: 26.96% Mismatches: 57
 Query Match: 13.78% Indels: 4
 DB: 1 Gaps: 1
 3-09-920-953-2 (1-598) x GLB_PARCA (1-116)
 QY 67 CTGTTTGATGACCTGGCGGCGCAGAGCATGAGCTGGCGGTGACACCTCTTACGAT 126
 Db 2 LeuPheGluGlnLeuGlyGlyGlnAlaAlaValGlnAlaValThrAlaGlnPheTyrAla 21
 QY 127 AAGTCTGCTGCTACCGGAGCTGCTGCTTCTTCTGAGTCCCTGGACATGCAAGAGCAG 186
 Db 22 AsnIleGlnAlaAspAlaThrValAlaThrPhePheAsnGlyIleAspMetProAsnGln 41
 QY 187 AACATGACGAGCTCAAGTTTCATGAGCTTCTGTTGGCGGAGCAGACCAATACAAAGGC 246
 Db 42 ThrAsnLysThrAlaAlaPheLeuCysAlaAlaLeuGlyGlyProAsnAlaTTrpThrGly 61
 QY 247 CGAAGCATGTACGACGACACACGCCCATCTGTGTCAAGGCGCCAGCGCTGGACACCGCCAC 306
 Db 62 ArgAsnLeuLysGluValHisAlaAsnMet-----GlyValSerAsnAlaGln 77
 QY 307 TTTGACAGATCAAGCATGACCTTGGAGACGCTGCAAGAGATGGCGCTCAAGCAGGAT 366
 Db 78 PheThrThrValIleGlyHisLeuArgSerAlaLeuThrGlyAlaGlyValAlaAla 97
 QY 367 GTGATCCAGCACCCCGCGGAGTGGTGAGTCCACCGCCGACGAA 411

Db 98 LeuValGluGlnThrValAlaValAlaGluThrValArgGlyAsp 112
 RESULT 10
 IE18_PRVKA STANDARD: PRT: 1446 AA.
 ID IE18_PRVKA
 AC P33479;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Immediate-early protein IE180.
 GN IE.
 OS Pseudorabies virus (strain Kaplan) (PRV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxID=33703;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91021039; PubMed=2171211;
 RA Vitek C., Kozmik Z., Paces V., Schirm S., Schwyz M.;
 RT "Pseudorabies virus immediate-early gene overlaps with an oppositely
 RT oriented open reading frame: characterization of their promoter and
 RT enhancer regions.";
 RL Virology 179:365-377(1990).
 CC -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
 CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
 CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
 CC -!- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
 CC -!- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
 CC PHOSPHORYLATION.
 CC -!- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M34651; AAA47470.1;
 DR PIR; A45344; A45344.
 DR InterPro; IPR005205; Herpes_ICP4_C.
 DR InterPro; IPR005206; Herpes_ICP4_N.
 DR Pfam; PF03584; Herpes_ICP4_N; 1.
 DR Pfam; PF03585; Herpes_ICP4_C; 1.
 KW Early protein; Transcription regulation; Trans-acting factor;
 FT DNA-binding; Phosphorylation; Nuclear protein.
 FT DOMAIN 347 354 POLY-SER.
 FT DOMAIN 379 397 POLY-SER.
 SQ SEQUENCE 1446 AA; 148640 MW; 81F43A3DE3DDA068 CRC64;
 Alignment Scores:
 Pred. No.: 0.177 Length: 1446
 Score: 106.50 Matches: 67
 Percent Similarity: 31.07% Conservative: 20
 Best Local Similarity: 23.93% Mismatches: 51
 Query Match: 9.98% Indels: 142
 DB: 1 Gaps: 12
 US-09-920-953-2 (1-598) x IE18_PRVKA (1-1446)
 QY 23 CCAGCACACAGAGCGCGGAGCGGATGCGGCGCAAGAAGCTGTTTGTGATGACCTGG 82
 Db 879 ProlAProGluArgLysLysLysLys-ArgArgAla-----ProG1 892
 QY 83 CGCGCGCAGAACGATCAAGCTGCGGCTTGTACACCTTCTACGATAAGGTCTCGCTGACC 142
 Db 892 yAlaArgArg-----Pr 896
 QY 143 CGGA-----GCTGCTGCCCTTCTTCGAGTCCCTCGACATGC 178
 Db 896 oGlyAspGlyGluGluAspGluGlyLeuSerGlyAlaAlaLeuArgGlyAspGlyHisG1 916


```
OY 179 AAGACGACGAGTGAAGCAGGTCAAGTTTCATGAGCTTCCTGTTGGCGGACGACGACCAAT 238
Db 916 yHisArgAspGluGluAsp----- 923
OY 239 ACAAGGCGCAGCATGTACGACGACACACGCCCATCTGTCGAAGGCGCCACGCGCTGGACC 298
Db 924 -ArgGlyProArgArgGlyArgSerLeuGlyLeuGlyProAlaProAspProAlaP 943
OY 299 ACCGCCACTTGCACAGATCAGCAGTACCTGGAG----- 334
Db 943 o--AlaLeuLeuSerSerSerSerSerGluAspArgLeuArgArgProLeu 962
OY 335 -----AGACGC----- 340
Db 963 GlyProMetProGluHisProAlaProAspGlyGlyPheArgArgValProAlaGlyGlu 982
OY 341 -----TCCAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCGC----- 382
Db 983 ThrHisThrProArgProSerGluAlaAlaLeuAlaAlaTyrcysProProGluValAla 1002
OY 383 -----CCGAGTGGTGGAGTCCACCCGCGACGAAATTG 415
Db 1003 ArgAlaLeuValAspGlnGluValPheProGluLeuTrpArgProAlaLeuThrPhe--- 1021
OY 416 ACTNCCCAACAACGCGCACCACTGATTTTCAATTAACCCACCCCGCTGAGCGCT 475
Db 1022 -----AspProAlaAlaLeuAla 1027
OY 476 CATTCCATCGATTGAGCGGGAGCGCCAGTT----- 508
Db 1028 His---IleAlaAlaArgArgGlyAlaProLeuArgArgAlaAlaTrpMetArgGln 1046
OY 508 ----- 508
Db 1047 IleAlaAspProGluAspValArgValValLeuTyrrAspProLeuProHisGluGlu 1066
OY 509 -----GCCAGCGCGCCAGGCGCCAGCAGCCTCA-----AATCGT 547
Db 1067 LeuGlyAlaGluProAlaGluGlyAlaProArgProAlaTrpAspProArgArgGlyGly 1086
OY 548 TTGCCAGCCCTTCTGCTGATTAAGAGCCATCAG-----CCATTTTGGCAC 592
Db 1087 LeuSerAlaLeuAlaAlaPheAlaHisArgLeuGlyCysThrProAspSerHis 1104
RESULT 11
ID HB9_HUMAN STANDARD; PRT: 401 AA.
AC P50219;
OY 01-OCT-1996 (Rel. 34, Created)
OY 01-OCT-1996 (Rel. 34, Last sequence update)
OY 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein HB9.
GN HLXB9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=94327547; PubMed=7914194;
RA Harrison K.A., Druey K.M., Deguchi Y., Tuscano J.M., Kehrl J.H.;
RT "A novel human homeobox gene distantly related to proboscipedia is
RL J. Biol. Chem. 269:19968-19975(1994).
CC -!- FUNCTION: PUTATIVE TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LYMPHOID AND PANCREATIC TISSUES.
CC -!- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U07664; AAB60647.1; JOINED.
DR EMBL; U07663; AAB60647.1; JOINED.
DR HSSP; P14653; 1B72.
DR TRANSFAC; T03420; -.
DR Genew; HGNC:4979; HLXB9.
DR MIM; 142394; -.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR Homeobox; DNA-binding; Nuclear protein; Transcription regulation.
KW DOMAIN 39 48 POLY-GLY.
FT DOMAIN 97 111 POLY-GLY.
FT DOMAIN 120 135 POLY-ALA.
FT DOMAIN 169 177 POLY-ALA.
FT DNA_BIND 242 301 HOMEBOX.
FT DOMAIN 316 325 POLY-GLY.
SQ SEQUENCE 401 AA; 40932 MW; 0006AED71D594FE CRC64;

Alignment Scores:
Pred. No.: 2.51 Length: 401
Score: 92.50 Matches: 44
Percent Similarity: 34.97% Conservative: 6
Best Local Similarity: 30.77% Mismatches: 39
Query Match: 8.24% Indels: 54
DB: 1 Gaps: 9

US-09-920-953-2 (1-598) x HB9_HUMAN (1-401)
OY 436 GTCCGCGAGTGTGGGNAAGTCAAATTCGTCGCGGGTGACT-----CCACCACCTC 386
Db 9 IleGluProCysTrpArg-----TrpThrProHisGluProProLeu 22
OY 385 CGCGCGGCTGCTGGATCACATCTCTGTGACGCCCATCTCTTCAGCGCTCTCTCCAAGT 326
Db 23 AlaGluArgAlaLeuAlaLysValThr---SerProValProAlaSer-----Gly 39
OY 325 ACTGCTTGATCTGTCAAAGTGGCGGTGTCAGCGCGTGGCCCTTGACCAGATGGCGCT 266
Db 40 Thr-----GlyGlyGlyGlyGlyGlyGly-----AlaSerGlyGly 52
OY 265 GTGCGTGTACATGCTTCGCGCCCTTCTATTGCTCTCTCCGCCAACACGCAAGCTCATGA 206
Db 53 ThrSerGlySerCysSer----- 58
OY 205 ACTTGACCTGCTTCATCTTCTGCTCTGCTGTCAGGAGCTCGAAGAGGGCAGCAGCT 146
Db 59 -----ProAlaSerSerGluProProAlaAlaProAlaAspArgLeuArgAlaGluSer 76
OY 145 CGGGGTACCCAGCACCTTATCGTAGAAGGTGTCAACCGCCAGCTTTCATGCTTCGCGC 86
Db 77 ProSerProProArg-----LeuLeu--- 83
OY 85 CGCCGAGGTATCAAAACAGCTTCTTTCGCCCGCAGCATCCCGCTTCGCGCGCTCTCTGTGCTG 26
Db 84 -----AlaAlaHisCysAlaLeuLeuProLysProGlyPheLeu-GlyAl 98
OY 25 TGTGTGGC 19
Db 98 aclygly 100
RESULT 12
ID UPBN_MOUSE STANDARD; PRT: 566 AA.
AC Q90ZL6;
```

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 23 (EC 3.1.2.15) (Ubiquitin
 DE thiolesterase 23) (Ubiquitin-specific processing protease 23)
 DE (Deubiquitinating enzyme 23).
 GN USP23.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20246307; PubMed=10786635;
 RA Smith T.S., Southan C.;
 RT "Sequencing, tissue distribution and chromosomal assignment of a novel
 RT ubiquitin-specific protease USP23.";
 RL Biochim. Biophys. Acta 1490:184-188(2000).
 CC -!- FUNCTION: Capable of removing ubiquitin from ubiquitinated
 CC proteins; also capable of removing NEDD8 from NEDD8 conjugates but
 CC has no effect on Sentrin-1 conjugates (By similarity).
 C -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF177759; RAD54322.1;
 DR MEROPS: C19.034;
 DR MGD: MGI:1353665; Usp23.
 DR InterPro: IPR001394; UCH-2.
 DR Pfam: PF00442; UCH-1; 1.
 DR Pfam: PF00443; UCH-2; 1.
 DR PROSITE: PS00972; UCH_2.1; 1.
 DR PROSITE: PS00973; UCH_2.2; 1.
 DR PROSITE: PS02335; UCH_2.3; 1.
 KW Ubiquitin conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT-SITE 221 221 BY SIMILARITY.
 FT ACT-SITE 511 511 BY SIMILARITY.
 FT ACT-SITE 519 519 BY SIMILARITY.
 SQ SEQUENCE 566 AA; 62672 MW; F2E1828E9BDB1AFD CRC64;

 Alignment Scores:
 red. No.: 2.59 Length: 566
 score: 92.50 Matches: 44
 Percent Similarity: 35.00% Conservative: 19
 Best Local Similarity: 24.44% Mismatches: 68
 Query Match: 8.67% Indels: 50
 DB: 1 Gaps: 7

 US-09-920-953-2 (1-598) x UBPN_MOUSE (1-566)
 QY 141 CCGGAGCTGCTGCTTCTTCGAGTCCTGACATGCAAGACGAGATGAGCAGGT 200
 Db 21 ProArgValGlyAlaLysIleProPheProProArgAlaArgSerLysGlu----- 37
 QY 201 CAAGTTTCAGAGCTGCTGTTTGGCGGAGCAGACCAATCAAGGGCGGAGGATGACGA 260
 Db 38 -----ArgArgAsnProValProGlyProAsnSerMetLeu 49
 QY 261 CGCACAGCCCATCTGCTCAAGGGCCGCTGGACCCGACCCATCTTCACAGATCAA 320
 Db 50 ArgProLeuPro-----ProArgProGlyProPro-----AspGlu 61
 QY 321 CGAGTACCTGGAGAGACGCTGCAAGAGATGGCGGTCAAGCAGGATGTATCCAGC---- 376
 Db 62 ArgLeuLysLysLeuGluLeuGlyArg---GlyArgThrSerGlySerArgProArgGly 80

QY 377 -----ACGCGCGCGAGTGTGTGAGTCCACCCCG----- 406
 Db 81 ProLeuArgAlaAspHisGlyValProLeuProGlySerProProAlaValAlaLeu 100
 QY 407 -----ACGAATTGTGACTTNCACAACTCCGCGACCCCACTGATTTTCAT 451
 Db 101 ProLeuProSerArgThrAsnLeuAlaArgSerLysSerValSerGlyAspLeuArg 120
 QY 452 TAACCAACCCAGCTGAGCGCTCATTCATCGATTTTTCAGCGGGGAGCGCGAGTGC 511
 Db 121 PrometGlyIleAlaLeuGlyGlyHisArgGlyAlaGlyGluLeuGlyAlaLeuSer 140
 QY 512 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 550
 Db 141 ArgLeuAlaLeuArgProGluProProThrLeuArgSerThrSerLeuArgArgLeu 160
 QY 551 -----CCAGCCCTTGTGTCATTCAGAGCGCATCAGGCATTTTCGCAC 592
 Db 161 GlyGlyPheProGlyProProThrLeuLeuSerIleArgThrGluProProThrSerHis 180
 RESULT 13
 ICP4_HSV11
 ID ICP4_HSV11 STANDARD; PRT; 1298 AA.
 AC P08392;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Trans-acting transcriptional protein ICP4 (transcriptional activator
 DE IE175) (Alpha-4 protein).
 GN ICP4 OR IE175 OR RS1.
 OS Herpes simplex virus (type 1 / strain 17).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Simplexvirus.
 OX NCBI_TaxID=10299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88274327; PubMed=2839594;
 RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
 RA McNab D., Perry L.J., Scott J.E., Taylor P.;
 RT "The complete DNA sequence of the long unique region in the genome of
 RT herpes simplex virus type 1.";
 RL J. Gen. Virol. 69:1531-1574(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86148504; PubMed=3005980;
 RA McGeoch D.J., Dolan A., Donald S., Brauer D.H.K.;
 RT "Complete DNA sequence of the short repeat region in the genome of
 RT herpes simplex virus type 1.";
 RL Nucleic Acids Res. 14:1727-1745(1986).
 RN [3]
 RP DNA-BINDING DOMAIN.
 RX MEDLINE=90174974; PubMed=2155403;
 RA Wu C.-L., Wilcox K.W.;
 RT "Codons 262 to 490 from the herpes simplex virus ICP4 gene are
 RT sufficient to encode a sequence-specific DNA binding protein.";
 RL Nucleic Acids Res. 18:531-538(1990).
 RN [4]
 RP INFLUENCE OF PHOSPHORYLATION ON FUNCTION.
 RX MEDLINE=91122047; PubMed=1846804;
 RA Papavassiliou A.G., Wilcox K.W., Silverstein S.J.;
 RT "The interaction of ICP4 with cell/infectious-cell factors and its
 RT state of phosphorylation modulate differential recognition of leader
 RT sequences in herpes simplex virus DNA.";
 RL EMBO J. 10:397-406(1991).
 CC -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
 CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
 CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS. BINDING
 CC OF ICP4 TO ALPHA GENES CAUSES REPRESSION OF THEIR EXPRESSION.
 CC ICP4 ALSO ACTS AS A TRANSCRIPTIONAL ACTIVATOR OF BETA AND GAMMA
 CC GENES. ICP4 BINDS WITH HIGH AFFINITY TO THE SEQUENCE 5'-ATCGTC-3'.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.

Db 1076 Met-----AspAsnLeuLysAsnSerArgLeuAlaThrAlaGluProVal 1090
 QY 378 CGCGCGGAGTGGTGGAGTCCACCGCGAGCAATTTGACTTNCCTCAACAACTCCGCACC 437
 Db 1091 SerProHisGluAsnLeuSerHisAlaGlyLeuProGlnSerProAlaLysMetGlySer 1110
 QY 438 CAACTGATTTTAACTCAACCAACCCAGCTGAGCGCTCATTCATCGATTTTGACGGGG 497
 Db 1111 SerThrAspProAlaGlyProThrProAlaThrAlaAlaAsnProGlnAsnSerThrAla 1130
 QY 498 GAGCGCCAGTTGCCGAGCGCGCCAGGGGCC 530
 Db 1131 SerArgArgThrProAsnAsnProGlyAsnPro 1141

RESULT 15

YT35_STRFR
 ID YT35_STRFR STANDARD: PRT: 348 AA.
 AC P20186;
 DT 01-FEB-1991 (Rel. 17, Created)
 RT 01-FEB-1991 (Rel. 17, Last sequence update)
 RL 16-OCT-2001 (Rel. 40, Last annotation update)
 OS Hypothetical 35.5 kDa protein in transposon TN4556.
 OC Streptomyces fradiae.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1906;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-TN4556;
 RX MEDLINE=90185236; PubMed=2155856;
 RA Siemieniak D.R., Slightom J.L., Chung S.T.;
 RT "Nucleotide sequence of Streptomyces fradiae transposable element
 TN4556: a class-II transposon related to Tn3.";
 RL Gene 86:1-9(1990).
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: M29297; AAA8561.1; -
 DR PIR: JQ0431; JQ0431.
 KW Hypothetical protein; Transposable element.
 SQ SEQUENCE 348 AA; 35520 MW; 3BDF5D83ABBB92A CRC64;

Alignment Scores:

ed. No.:	4.06	Length:	348
Score:	90.00	Matches:	38
Percent Similarity:	43.24%	Conservative:	10
Best Local Similarity:	34.23%	Mismatches:	48
Query Match:	8.43%	Indels:	15
DB:	1	Gaps:	4

US-09-920-953-2 (1-598) x YT35_STRFR (1-348)

QY 116 CTTTCTACGATAGGTGCTGCTGACCGGAGCTGCTCTTCGAGTCCCTGGACA 175
 Db 212 ProLeuAlaAlaArgCysSerAlaSerArgAlaGlyCysProThrAlaAlaGlySerLeu 231
 QY 176 TGCAGAGCAGACAGA-----TGAAGCAGGTCAAGTTCATGAGCTTCGTGTTGGCG 226
 Db 232 LeuProAlaProArgProProAlaSerSerAlaSerProGlnAlaAlaProAla 251
 QY 227 GAGCAGACCAATCAAGCGCGGAGCATGTACGACGACAGCCCATCTCGTCAAGGGCC 286
 Db 252 AlaProSerAlaThrArgLeu-Pro-----ArgArgThrProSerAlaProArgPr 269
 QY 287 A-----CGCCTGGACCGCCACTTTGACAAGATCAAGCAGTACCTTGGAGAGAC-- 338
 Db 269 oSerSerArgProAlaArgProProLleProAlaAlaArgProProArgThrPr 289

QY 339 -----GCTCAAGAGATGGCGCTCAAGCAGGATGTGATCCAGCACG 379
 Db 289 OGlyThrProArgProAlaAlaAlaArgAlaArgAlaProAlaGlyCysSerProAlaAr 309
 QY 380 CCGCCGCGAGTGTGGAGTCCACCCCGCGAGCA 410
 Db 309 gArgThrProSerAlaProThrAspArg 319

Search completed: April 16, 2003, 13:01:03
 Job time : 18 secs


```

Query Match: 23.90% Indels: 0
DB: 2 Gaps: 0

US-09-920-953-2 (1-598) x Q8RT58 (1-124)

QY 67 CTGTTTGATGACCTGGCGCGCCAGAGGCATGAAGCTGGCGTTGACACCTTCTACGAT 126
DB 4 LeuTyrGluLysLeuGlyGlyAlaAlaAlaValAspLeuAlaValGluLysPheTyrGly 23
QY 127 AAGTGCTGGCTGACCGGAGCTGCTTCTTCGAGTCCCTGCACATGCAAGAGCAG 186
DB 24 LysValLeuAlaAspGluArgValAlaAsnArgPhePheValAsnThrAspMetAlaLysGln 43
QY 187 AGATGACAGCGTCAAGTTTCATGAGCTTCGTGTTGGCGGAGCAGACCAATACAGGCG 246
DB 44 LysGlnHisGlnLysAspPheMetThrTyrAlaPheGlyGlyThrAspArgPheProGly 63
QY 247 CGAAGCATGACGACGACCCCATCTGCTGTCAGAGCCGCGCTGACCGCTGACCCGCGCAC 306
DB 64 ArgSerMetArgAlaAlaHisGlnAspLeuValGluAlaValGluAlaValThrAspValHis 83
QY 307 TTTGACAAGATCAAGCAGTACCTGGAGAGCTGCTGCAAGAGATGGCGTCAAGCAGGAT 366
DB 84 PheAspAlaIleAlaGluAsnLeuValLeuThrLeuGlnGluLeuAlaValSerGlnAsp 103
QY 367 GTGATCCAGCAGCGCGCGAGTGTGGAGTCCACCGCGGAGCAATTTGACTTNCACAC 426
DB 104 LeuIleAspGluValValThrIleValGlySerValGlnHisArgAsnAspValLeuAsn 123

RESULT 2
Q27224
ID Q27224 PRELIMINARY; PRT: 117 AA.
AC Q27224;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HEMOGLOBIN
OS Paramesidium jenningsi, and
OS Paramesidium triaurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramesidium.
OX NCBI_TaxID=44029, 44031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96038820; Pubmed=7578257;
RA Yamauchi K., Tada H., Usuki I.;
RT "Structure and evolution of Paramesidium hemoglobin genes.";
RL Biochim. Biophys. Acta 1264:53-62(1995).
DE EMBL: D49688; BAA08540.1;
EMBL: D49688; BAA08539.1;
InterPro: IPR001486; Bac_globin.
DR Pfam: PF01152; Bac_globin; 1.
DR ProDom: PD004840; Bac_globin; 1.
DR PROSITE: PS01213; GLOBIN_FAM.2; 1.
SQ SEQUENCE 117 AA; 12090 MW; 51A8C7F7E1E5106B CRC64;

Alignment Scores:
Pred. No.: 2.34e-05 Length: 117
Score: 149.00 Matches: 31
Percent Similarity: 46.96% Conservative: 23
Best Local Similarity: 26.96% Mismatches: 57
Query Match: 13.96% Indels: 4
DB: 5 Gaps: 1

US-09-920-953-2 (1-598) x Q27224 (1-117)

QY 67 CTGTTTGATGACCTGGCGCGCCAGAGGCATGAAGCTGGCGTTGACACCTTCTACGAT 126
DB 3 LeuPheGluGlnLeuGlyGlyGluAlaAlaValThrAlaValThrGlnPheTyrAla 22
QY 127 AAGTGCTGGCTGACCGGAGCTGCTTCTTCGAGTCCCTGCACATGCAAGAGCAG 186
DB 23 AsnIleGlnAlaAspAlaThrValAlaAsnPhePheAsnGlyIleAsnMetProThrGln 42
QY 187 AGATGACAGCGTCAAGTTTCATGAGCTTCGTGTTGGCGGAGCAGACCAATACAGGCG 246
DB 43 ThrAspLysThrAlaAlaPheLeuCysAlaAlaLeuGlyGlyProAsnAlaTrpAlaGly 62
QY 247 CGAAGCATGACGACGACCCCATCTGCTGTCAGAGCCGCGCTGACCGCTGACCCGCGCAC 306
DB 63 ArgAsnLeuLysGluValHisAlaAsnMet-----GlyValThrAsnAlaGln 78
QY 307 TTTGACAAGATCAAGCAGTACCTGGAGAGCTGCTGCAAGAGATGGCGTCAAGCAGGAT 366
DB 79 PheThrThrValIleGlyHisLeuArgSerAlaLeuThrGlyAlaGlyValAlaAla 98
QY 367 GTGATCCAGCAGCGCGCGAGTGTGGAGTCCACCGCGGAGCA 411

US-09-920-953-2 (1-598) x Q27213 (1-118)

QY 67 CTGTTTGATGACCTGGCGCGCCAGAGGCATGAAGCTGGCGTTGACACCTTCTACGAT 126
DB 3 LeuPheGluGlnLeuGlyGlyGluAlaAlaValThrAlaValThrGlnPheTyrAla 22
QY 127 AAGTGCTGGCTGACCGGAGCTGCTGCTTCTTCGAGTCCCTGCACATGCAAGAGCAG 186
DB 23 AsnIleAlaAlaAspAlaThrValAlaAsnPhePheAsnGlyIleAsnMetProThrGln 42
QY 187 AGATGACAGCGTCAAGTTTCATGAGCTTCGTGTTGGCGGAGCAGACCAATACAGGCG 246
DB 43 ThrAspLysThrAlaAlaPheLeuCysAlaAlaLeuGlyGlyProAsnAlaTrpAlaGly 62
QY 247 CGAAGCATGACGACGACCCCATCTGCTGTCAGAGCCGCGCTGACCGCTGACCCGCGCAC 306
DB 63 ArgAsnLeuLysGluValHisAlaAsnMet-----GlyValThrAsnAlaGln 78
QY 307 TTTGACAAGATCAAGCAGTACCTGGAGAGCTGCTGCAAGAGATGGCGTCAAGCAGGAT 366
DB 79 PheThrThrValIleGlyHisLeuArgSerAlaLeuThrGlyAlaGlyValAlaAla 98
QY 367 GTGATCCAGCAGCGCGCGAGTGTGGAGTCCACCGCGGAGCA 411

Alignment Scores:
Pred. No.: 0.000243 Length: 118
Score: 138.00 Matches: 30
Percent Similarity: 46.09% Conservative: 23
Best Local Similarity: 26.09% Mismatches: 58
Query Match: 12.93% Indels: 4
DB: 5 Gaps: 1

```

Db	99	LeuVaIGluGlnThrValAlaVaAlaGluThrValArgClyAsp	113
RESULT 4			
Q9GND3			
ID	Q9GND3	PRELIMINARY;	PRT; 116 AA.
AC	Q9GND3;		
DT	01-MAR-2001	(TREMBLrel. 16, Created)	
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)	
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)	
DE	Hemoglobin.		
OS	Paramesium caudatum.		
OC	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; P		
OC	Paramesium.		
OX	NCBI_TaxID:5885;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Yamauchi K.;		
RT	"Heterogeneous Compositions in Paramesium caudatum Mac		
RT	Resulting from Various Patterns of Excision of Internal		
RT	Sequence.";		
RT	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases		
..	[2]		
RN	SEQUENCE FROM N.A.		
RA	Yamauchi K.;		
RT	"Heterogeneous compositions in Paramesium caudatum mac		
RT	resulting from various patterns of excision of interna		
RT	sequence.";		
RT	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases		
DR	EMBL; AB023083; BAB20310.1;		
DR	EMBL; AB023078; BAB20307.1;		
DR	EMBL; AB023080; BAB20308.1;		
DR	EMBL; AB023082; BAB20309.1;		
DR	InterPro; IPR001486; Bac-globin.		
DR	Pfam; PF01152; Bac-globin; 1.		
DR	ProDom; PD004840; Bac-globin; 1.		
DR	SEQUENCE 116 AA; 11882 MW; 63B1D24F28C384EE CRC64;		

Alignment Scores:	
Pred. No.:	0.00964
Score:	131.50
Length:	116
Percent Similarity:	46.09%
Best Local Similarity:	26.09%
Conservative:	23
Mismatches:	57
Query Match:	12.32%
Indels:	5
DB:	5
Gaps:	2
Mismatches:	23
Conservative:	57
Indels:	5
Gaps:	2

US-09-920-953-2 (1-598) x 09GND3 (1-116)

```

67 CTGTTGATGACCTGGCGCGGCAGGAAGCATGAAGCTGCGGCTTGACACCTTCTACGAT 126
      |||||::: |||||::: |||||::: |||||:::
3  LeuPheGluGlnLeuGlyGlyGlnAlaAlaValGlnAlaValThrAlaGlnPheYrAla 22
      |||||::: |||||::: |||||::: |||||:::
127 AAGTGCTGCTCACCGGAGCTGCTGCCCTTCTCGAGTCCCTCGACATGCAAGGACG 186
      ::: |||||::: |||||::: |||||::: |||||:::
23  AsnIleGlnAlaAspAlaThrValAlaThrPhePheAsnGlyIleAspMetProAsnGln 42
      ::: |||||::: |||||::: |||||::: |||||:::
187 AAGATGAAGCAGGCTCAAGTTCATGAGCTCGTGTGGCGGAGACCAATCAAGGCG 246
      ||| |||||::: |||||::: |||||::: |||||:::
43  ThrAsnLysThrAlaAlaPheLeuCysAlaAlaLeuGlyGlyProAsnAlaTrpThrGly 62
      |||||::: |||||::: |||||::: |||||:::
247 CGAAGCATGTACGACGACACGCCCATCTGGTCAAGGGCAGCGGCTGCACCCGCCAC 306
      ::: |||||::: |||||::: |||||::: |||||:::
63  ---AsnLeuLysGluValHisAlaAsnMet-----GlyValSerAsnAlaGln 77
      |||||::: |||||::: |||||::: |||||:::
307 TTTCACAAGATCAAGCATGCTACCTTGGAGAGACGCTGCACAGATGGGCGTCAAGCAGGAT 366
      ||| ::: |||||::: |||||::: |||||:::
78  PheThrThrValIleGlyHisLeuArgSerAlaLeuThrGlyAlaGlyValAlaAla 97
      |||||::: |||||::: |||||::: |||||:::
367 GTGATCCAGCACCGCGCGGAGTGGTGGAGTCCACCCCGGAGGAA 411
      ::: |||||::: |||||::: |||||:::
98  LeuValGluGlnThrValAlaValAlaGluThrValArgGlyAsp 112
      |||||::: |||||::: |||||::: |||||:::

```

```

Q9H7T3
ID Q9H7T3 PRELIMINARY; PRT; 257 AA.
AC Q9H7T3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CDNA FLJ14280 fis, clone PLACE1005584, weakly similar to trans-acting
DE transcriptional protein ICPO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=PLACENTA.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wakatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Yanomiya K., Iwayanagi T.,
RT "NEDO human cDNA sequencing project.";
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK024342; BAB14892.1;
SQ SEQUENCE 257 AA; 26269 MW; B914A57EA54DF418 CRC64;

Alignment Scores:
Pred. No.: 0.025 Length: 257
Score: 117.00 Matches: 60
Percent Similarity: 38.10% Conservative: 20
Best Local Similarity: 28.57% Mismatches: 72
Query Match: 10.97% Indels: 58
DB: 4

US-09-920-953-2 (1-598) x Q9H7T3 (1-257)

```

Qy	26	GCACCACAGAGCGCGGAAGCGGGATGCGGGCGCAAGAAGCTGTTTGTATGACCTGGCGG	85
		:::	
Db	34	AlaProSerLysAsnSerSerArg---LeuGlyArgProCysMetCysThr---Ala	51
Qy	86	GCACAGAAGGCATGAAGCTGGCGGTTCACACCTTCACATAAAGTGTGGCTGACCCGG	145
		::::	
Db	52	GlyArgArg-----ProAsnArgAlaSerGlyArgArgArgArg	64
Qy	146	AGTGTCTGCCCTTCTTCGACT-----CCCTGGCACATGC-----AAGAGCAGAAGATCA	193
		:::	
Db	65	SerCysSerProAlaProThrTppProProLeuCysCysTyProGlnSerArgProThr	84
Qy	194	AGCAGGTCAAAGTTCATGAGCTTCGTGTGGCGGAC-----AGACCA-----	236
		::: ::::	
Db	85	AlaSerAlaAlaGlyProGlyAlaCysMet-ArgAlaSerGlyArgProHisGlyAsnTh	104
Qy	237	-----ATCAAGGGCGGAAGCATGTACACGCCACACGCCCATCTGCTCAAGGG-	284
Db	104	rThrAlaSerThrAlaProProArgHisProArgProArgProGlyGlyProAlaLe	124
Qy	285	-----CCAGG-----CCTGGACCACCGCCCACTTTGCACAAGT-----	317
Db	124	uArgProThrProArgProCysAlaGlyProAlaProProProAlaSerArgAspCys	144
Qy	318	-----CAAGCAGTACTTTGGAGACCGGTGCAAGAGATGGCGGTCAACAGCAGGA-	367
		:::	
Db	144	gCysArgArgProArgArgTrpProArgAlaGlyArgArgGlyArgArgAlaGlyAlaC	164
Qy	368	TGATCCAGCAGCGCGCGGAGTGGTGAGTCCACCGCGACGAATTTGACTTNCACAAC	427
Db	164	sLysProSer-CysAlaGlyAlaAlaLrpSerAlaArgGly-----	177
Qy	428	ACTCGGCACCCACTGATTTTCATTAAACCCACCCGAGCTGAGCGCTCATTCATCGAT	487
Db	178	-----AlaProLeuCysSerTyrArgThr	186

RESULT 5

QY 488 TTTGACGGGGGAGCCAGCTTCCGAGCGCCGAGCGCCGAGGAGCCCTCAAAATCG- 546
|||||
Db 186 erCysAlaGlySerCysGlyAlaArgThrAlaProThrProAlaProThrCysAlaSerP 206
|||||
QY 547 -----TTTGACGACCTTGTCTGC 564
|||||
Db 206 roSerAlaAlaAlaSerSerCysCys 214
|||||
RESULT 6
Q98N74 PRELIMINARY: PRT: 160 AA.
AC Q98N74: (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DE Hypothetical protein mlr0267.
GN MLR0267.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
NCBI_TaxID=381;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099; PubMed=11214968;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
Watanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,
Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
Takeuchi C., Yamada M., Tabata S.
RT *Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti.;
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002994; BAB47888.1;
DR InterPro; IPR001486; Bac_globin.
DR Pfam; PF01152; Bac_globin; 1.
DR ProDom; PD004840; Bac_globin; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 160 AA; 17777 MW; DBCFCE748FA4EB0B CRC64;

Alignment Scores:
Pred. No.: 0.0474 Length: 160
Score: 113.50 Matches: 40
Percent Similarity: 40.95% Conservative: 27
Best Local Similarity: 24.39% Mismatches: 70
Query Match: 10.64% Indels: 27
DB: 16 Gaps: 7

--09-920-953-2 (1-598) x Q98N74 (1-160)

QY 67 CTGTTTGATGACCTGGCGGCGCAGAGGCATGAAGCTGGCGGTTGACACCTTCTACGAT 126
|||||
Db 8 LeuTyrgluTrpAlaGlyGlySerAspAlaLeuAsnArgLeuThrGlnThrPheTyraSp 27
|||||
QY 127 AAGTGCTGCTGACCGGAGCTGCTGCTTCTCGACTCCCTGGACATCAAGAGCAG 186
|||||
Db 28 LysValAlaLysAspProValGlyProValPheLysAlaMetSer---ProAspHis 46
|||||
QY 187 AAGTAAGACGAGGTCAAGTTCATGAGCTTCGTTGGCGGAGCAGCAATACAAGGC 246
|||||
Db 47 ProSerHisValAlaAlaPheIleGlyGluValPheGlyGlyProLysThrTySerGlu 66
|||||
QY 247 CGAAGCATGTACGAGCAGCCCATCTGCTGAAGGCGGCGCTGGACCCAGCCGAC 306
|||||
Db 67 Lys-----PheGlyGlyHisArgGluMetValMetHisLysLeuGlyLysHisLeuThr 84
|||||
QY 307 TTGCACAGATCAACGACTAC-----CTTGAGAGAGCGCTGCAAGATGGCGTC 357
|||||
Db 85 GluGluGlnArgArgTrpIleAsnLeuLeuAlaAlaAspGluValGlyLeu 104
|||||
QY 358 AAGCAGGATGTGATCCAGCAGCGGCC-----GGAGTGGTGGAG-----TCC 399
|||||
Db 105 ProAspAspProGluPheArgSerAlaPheMetGlyTyrrValGluTrpGlySerArgLeu 124
|||||

Alignment Scores:
Pred. No.: 0.0885 Length: 640
Score: 112.00 Matches: 56
Percent Similarity: 33.94% Conservative: 18
Best Local Similarity: 25.69% Mismatches: 59
Query Match: 9.98% Indels: 85
DB: 10 Gaps: 17

US-09-920-953-2 (1-598) x Q96397 (1-640)

QY 590 GCGAAATGCTGATGCTCTTCAATGCAGCAGGCTGGCAACGATTTTCGAGGCTCCT 531
|||||
Db 446 AlaArgValAlaAlaGlyThrSer-----ArgAlaAla 456
|||||
QY 530 GGGCCCTGCGGCGCTGCGCAACTGCGCTCCCGCTCAAAATCGATGAATGAGCGC 471
|||||
Db 457 GlyValProArgArg-----LeuGlnArgArgTTPArgArgArg 469
|||||
QY 470 TCAGGCTGGGTTGGGTTAATAAATCAGTTGGGTCGCGCAGTTGTTGGGNAAGTCAAT 411
|||||
Db 470 ---GlyArgGlyTrp-----ArgArgArgValArgArgArgGly 482
|||||
QY 410 TCGCGCGGCTGGGACTCCACCACCTCCGCGCGCTGGATCACATCTGCTTG- 357
|||||
Db 483 AlaGlyArgAlaValCysThrAlaGlyArgCysCysTrpMetThr---CysLeuProMet 501
|||||
QY 356 -----ACGCCCATC---TCTTGGCAGC 339
|||||
Db 502 TrpGlySerGlyGlyThrTrpTrpArgProLeuMetThrProSerArgThrCysAla 521
|||||
QY 338 GTCTCTCAAGGACTGCTGATCTTGTCAAGTGG-----CGTGGTCCAGCCCGTGG 285
|||||
Db 522 CysLeuProThrProCys-----CysSerArgTrpLeuArgArgTrpGlyCysGlyTrp 539
|||||

NCBI_TaxID=86665;

BULLIS K.C.; BUSAM D.A.; BUTLER H.,
RA Cherry J.M.; Cawley S.; Dahlke C.

BUTLER K.C., BUTLER H., CADIEU E., CENTER A., CHANDRA I.,
CHERRY J.M., CAWLEY S., DAHLKE C., DAVENPORT I., DAVIES P.

Db	531	-----			::: :::	SerThrThrSerThrThrProProSerThrGlu	543
Qy	413	TTGACTTNCACAACTGGCCAC		::: :::	-----CCAAGTATTTTCAT	451	
Db	544	ServAlserProAnsnSerAsnAlaAsnThrAsnAlaThrpProAlaPOTyrHis		::: :::	-----	563	
Qy	452	TAAACCAACCACCGCTGAGCGTCATCCTCATCGATTGTGAGCGGGAGCG		::: :::	-----	502	
Db	564	--ProArgProGlyIleValLeuAspProGluPhelyysProGlyIlyArgProArg		::: :::	-----	582	
Qy	503	CCAGTTGCCAGGCGGCCAGGGGGCCAGGAGCGCTGCATAATCGT		::: :::	-----TTGCCA	553	
Db	583	ProProValGlnArg		::: :::	-----ProProAlaGlnGlnThrLeuProLeuPro	597	
Qy	554	GCCCTTGCTGCTATTGAAGAGCCCATCAGCCA		::: :::	-----	583	
Db	598	AlaValGlnProThrArgGlnHisLeuPro		::: :::	-----	607	
RESULT 12							
ID	O02661	PRELIMINARY;			PRT;	425 AA.	
OC	A02661;						
DT	01-JUL-1997 (TrEMBLrel. 04, Created)						
DT	01-JUL-1997 (TrEMBLrel. 04, Last sequence update)						
DE	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)						
DE	SCO-spondin (Fragment).						
OS	Bos taurus (Bovine).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;						
OX	Bovidae; Bovinae; Bos.						
OX	NCBI_TaxID=9913;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	TISSUE-SUBCOMMISSURAL ORGAN;						
RA	Gubron S.;						
RL	Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.						
DR	EMBL; Y08561; CAA69868.1; -						
DR	Inter-Pro; IPRO02919; TIL_Cysrich.						
DR	Inter-Pro; IPRO00884; TSPL.						
DR	Pfam; PF01826; TIL; 1.						
DR	Pfam; PF00950; tsp_1; 3.						
DR	SMART; SM00209; TSPL; 3.						
DR	PROSITE; PS0092; TSPL; 4.						
FT	NON_TER 1						
FT	NON_TER 425						
FT	NON_TER 425						
SQ	SEQUENCE 425 AA; 43721 MW; E63DD4AFC6DB395D CRC64;						
Alignment Scores:							
Pred. No.:	0.677	Length:	425				
Score:	102.00	Matches:	69				
Percent Similarity:	27.36%	Conservative:	12				
Best Local Similarity:	23.31%	Mismatches:	62				
Query Match:	9.09%	Indels:	153				
DB:	6	Gaps:	17				
US-09-920-953-2 (1-598) x O02661 (1-425)							
Qy	554	CTGCCAACCAATTTCAGGCTCTCTGG-		-----	-----	528	
Db	97	VaIcglygluLeuGlnGlnProGlyIyValProPheLeuLeuAspAsnCysSerArGys		-----	-----	116	
Qy	527	-----	CCCCCTGGCGCGCTCGGCAACTGGCGCTCCC		-----	495	
Db	117	VaIcysGluLysGlyAlaLeuLeuCysGluProGly-		-----	-----	131	
Qy	494	GCTCAAATCGATGGTAATGAGCGCTCAGGCTGGGGTTGGTTAATGAA-		-----	-----	444	
Db	132	VaIProcysGlyIrrpSerAlaTtpSerSerTriplyProCysAspargSerCysGlySer		-----	-----	151	
Qy	443	CAGTTGGTCGCGCAGTTGTTGGNNACTCAAT-		-----	-----	411	
Db	152	GlyLeuArgAlaArgPheArgserProSerAsnProAlaAlaSerGlyIyAlaPro		-----	-----	171	

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 16, 2003, 12:57:26 ; Search time 14.5 Seconds
(without alignments)
2426.884 Million cell updates/sec

Title: US-09-920-953-2
Perfect score: 1067
Sequence: 1 gcctgcctgcagacgcggc.....agccatttgcaccagcc 598

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

sarched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model
-Q=/cgn2_1/USPTO.spool/US09920953/runat_16042003_115021_8171/app_query.fasta_1.775
-DB=Issued_Patents_AA -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09920953 -ECGN_1_1_28=runat_16042003_115021_8171 -NCPU=6 -ICPU=3
-NO.XLPXY -NO.MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Issued_Patents_AA.*
- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	96	8.6	154	4	US-09-193-104-8
2	93	8.7	696	3	US-08-906-865-4
3	93	8.7	696	4	US-09-129-668-4
C 4	92.5	8.2	226	6	5198348-1
5	92.5	8.7	1298	2	US-08-690-473-2
6	92.5	8.7	1298	4	US-09-259-821A-2
7	92.5	8.7	1298	4	US-08-843-639-2
8	92	8.6	1213	4	US-09-413-814-79
C 9	89	7.9	154	4	US-09-193-104-9
C 10	89	7.9	154	4	US-09-193-104-10
C 11	89	8.3	355	4	US-08-483-533-41
12	89	8.3	355	4	US-09-283-471A-41
					Sequence 8, Appli
					Sequence 4, Appli
					Sequence 4, Appli
					Patent No. 5198348
					Sequence 2, Appli
					Sequence 2, Appli
					Sequence 2, Appli
					Sequence 79, Appli
					Sequence 9, Appli
					Sequence 10, Appli
					Sequence 41, Appli
					Sequence 41, Appli

13	89	8.3	355	5	PCT-US91-06532-3	Sequence 3, Appli
14	88.5	8.3	882	4	US-09-413-814-78	Sequence 78, Appli
C 15	88	7.8	397	5	PCT-US96-10602-6	Sequence 6, Appli
16	87.5	7.8	132	1	US-08-534-975-4	Sequence 4, Appli
C 17	87.5	7.8	132	3	US-09-129-855A-4	Sequence 4, Appli
C 18	87.5	7.8	132	4	US-09-247-714-4	Sequence 4, Appli
C 19	87.5	7.8	132	4	US-09-480-718-4	Sequence 4, Appli
C 20	87.5	7.8	201	4	US-09-220-528-116	Sequence 116, App
C 21	87.5	7.8	134	4	US-09-193-104-18	Sequence 18, Appli
C 22	87	7.8	395	6	5196194-18	Patent No. 5196194
C 23	87	7.8	882	4	US-09-413-814-78	Sequence 78, Appli
C 24	87	7.8	1024	4	US-09-562-737-42	Sequence 42, Appli
C 25	86.5	8.1	170	4	US-09-199-637A-299	Sequence 299, App
C 26	86.5	7.7	591	3	US-09-082-737-2	Sequence 2, Appli
C 27	86.5	7.7	154	4	US-09-193-104-15	Sequence 15, Appli
C 28	86	7.7	136	4	US-09-199-637A-357	Sequence 357, App
C 29	85.5	7.6	277	6	5164485-2	Patent No. 5164485
C 30	85.5	7.6	281	1	US-08-105-483-214	Sequence 214, App
C 31	85.5	7.6	281	1	US-08-709-209-214	Sequence 214, App
C 32	85.5	7.6	281	1	US-08-458-101-214	Sequence 214, App
C 33	85.5	7.6	389	1	US-08-105-483-216	Sequence 216, App
C 34	85.5	7.6	389	1	US-08-105-483-216	Sequence 216, App
C 35	85.5	7.6	389	1	US-08-709-209-216	Sequence 216, App
C 36	85.5	7.6	389	1	US-08-709-209-216	Sequence 216, App
C 37	85.5	7.6	389	1	US-08-458-101-216	Sequence 216, App
C 38	85.5	7.6	389	1	US-08-458-101-216	Sequence 216, App
C 39	85.5	7.6	389	1	US-08-458-101-219	Sequence 219, App
C 40	85	8.0	495	4	US-09-096-776B-2	Sequence 2, Appli
C 41	84.5	7.9	181	4	US-09-220-528-40	Sequence 40, Appli
C 42	84.5	7.9	220	4	US-09-220-528-26	Sequence 26, Appli
C 43	84.5	7.5	236	1	US-08-378-011A-1	Sequence 1, Appli
C 44	84.5	7.9	237	4	US-09-220-528-32	Sequence 32, Appli
C 45	84.5	7.9	323	1	US-08-591-969-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-193-104-8
; Sequence 8, Application US/09193104A
; Patent No. 6172193
; GENERAL INFORMATION:
; APPLICANT: Primi, Daniele
; APPLICANT: Fioralissi, Gianfranco
; APPLICANT: Palla, Mario
; TITLE OF INVENTION: Escape Mutant of the Surface Antigen of Hepatitis B
; TITLE OF INVENTION: Virus
; FILE REFERENCE: SBD1004US
; CURRENT APPLICATION NUMBER: US/09/193.104A
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: EP 97830635.5
; EARLIER FILING DATE: 1997-12-01
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Hepatitis B virus
US-09-193-104-8

Alignment Scores:
Pred. No.: 0.0403
Score: 96.00
Percent Similarity: 32.39%
Best Local Similarity: 23.30%
Query Match: 8.56%
DB: 4
Length: 154
Matches: 41
Conservative: 16
Mismatches: 46
Indels: 73
Gaps: 10

US-09-920-953-2 (1-598) x US-09-193-104-8 (1-154)

Oy 454 TGGGTTGGTAAATGAAATCAGTTGGTGGCAG-----TTCTGGGNAAGTCAAT 411
|||||
Db 2 Trp---TrpThrSerLeuAsnPhleuGlyGlyThrValCysLeuGlyGlnAnser 20
|||||

```
OY 410 TCGTCGGGTGACTCCACCACTCGGGGGGGTGTGTGTGATCATCATCTGCTGAGCGCC 351
    |||
Db 21 GlnSerProThrSerAsnHisSerProThrSerCys----- 32
OY 350 ATCTCTTGACGGTCTCTCCAAAGTACTGCTTGCATCTTGTCAAAGTGGCGGTGGTCCAGG 291
    |||
Db 33 -----ProProleCys----- 36
OY 290 CCGTGGCCCTTACCAGATGGCGTGTGCTGCTATCATGCTTCGGCCCTGTATTGGTCT 231
    |||
Db 37 -----ProGlyTyrArgPrpMetCysLeuArgPheLeuPheLeuPhe----- 52
OY 230 GCTCCGCCAACACGAGCTCATGACCTGACCTGCTTCATCTTC----- 186
    |||
Db 53 -----IleLeuLeuLeuCysLeuIlePheLeuLeuValLeuLeu 65
OY 185 -----TGCTCTTCATGTCACAGGACTCGAAGAGGCGCAGC 150
    |||
Db 66 AspTyrGlnGlyMetLeuProValCysProLeuIle-ProGlySerSerThrThrSerTh 85
    |||
OY 149 AGCTCCGGGTGACCCAGCACCTTATCGTAGAAGGTGTCAACGCCGAGCT----- 101
    |||
Db 85 rGlyPro-----CysArgThrCysThrThrProAlaGlnGlyThrSe 99
OY 100 -----TCATGCTTCTCGCGCCGAGGTATCAACAGCTTCTTGGCGCCGCGAT 51
    |||
Db 99 rMetTyrProSerCysCysThrLysProSerHisGlyAsnCysThrCys-----117
OY 50 CCGCTTCGGCGCTCT-----CTGTGGTGGTCTGG 23
    |||
Db 117 eProIleProSerSerTrpAlaPheGlyLysPheLeuTrpGluTrp 132

RESULT 2
US-08-906-865-4
; Sequence 4, Application US/08906865
; Patent No. 6040168
; GENERAL INFORMATION:
; APPLICANT: Greengard, Paul
; APPLICANT: Porton, Barbara
; APPLICANT: Kao, Hung-Teh
; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906.865
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1884
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 696 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: /desc - "Synapsin Ia"
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-906-865-4

Alignment Scores:
Pred. No.: 0.148 Length: 696
Score: 93.00 Matches: 45
Percent Similarity: 36.97% Conservative: 16
Best Local Similarity: 27.27% Mismatches: 64
Query Match: 8.72% Indels: 41
DB: 3 Gaps: 7

US-09-920-953-2 (1-598) x US-08-906-865-4 (1-696)
OY 81 GGGCGCGCCAGAGCATGAAGCTGGCGGTTGACACCTTCTACGATAAGGTGCTGGCTGA 140
    |||
Db 530 GlyArgGlnSerArgProValAlaGlyGly----- 539
OY 141 CCCGAGCTGCTGCCCTTCTTCGAGTCCCTGACATGCAAGACAGCAAGATGAACAGGT 200
    |||
Db 540 ProGlyAlaProProAlaAlaArgProAlaSerProSerProGlnArgGlnAlaGly 559
OY 201 CAAGTTCATGAGCTTCTGTTGGCGGAGCAGACCAACATACAGGCCCGCAAGCATGTACCA 260
    |||
Db 560 PROPRO-----GlnAlaThrArgGlnThrSerValSerGlyProAlaProProLys 576
OY 261 CGCACA---CGCCCATCTGTCAAGGGCCAGCGCTGGACCAACCCGACCTTTGACAAGAT 317
    |||
Db 577 AlaSerGlyAlaProProGlyGlnGlnArgGlnGlyProProGlnLysProProGly 596
OY 318 CAAGCAGTACCTTGGAGAGAGCTGCCAAGAGATGGG----- 353
    |||
Db 597 ProAlaGlyProThrArgGlnAlaSerGlnAlaGlyProValProArgThrGlyProPro 616
OY 354 -----CGTCAAGCAGGATGTGATCCAGCAGC-----CCGCCGAGTGG 391
    |||
Db 617 ThrThrGlnGlnProArgProSerGlnAspValProProAlaThrAlaAlaGlyGlyProPro 636
OY 392 TGGAGTCCACCCGCGAGCAATTTGACTTNCACCAACTGGG----- 433
    |||
Db 637 AlaGlnLysProSerGlnAspValProProAlaThrAlaAlaGlyGlyProPro 656
OY 434 CACCAACTGATTTCAT-----TAACCCCAACCCGAGCTGAGCGCTCAT 478
    |||
Db 657 HisProGlnLeuPheAsnLeuProGluProAlaProProArgProSerLeuSerGlnAsp 676
OY 479 TCCATCGATTTTGAG 493
    |||
Db 677 GluValLysAlaGlu 681

RESULT 3
US-09-129-668-4
; Sequence 4, Application US/09129668B
; Patent No. 6428010
; GENERAL INFORMATION:
; APPLICANT: Greengard, Paul
; APPLICANT: Porton, Barbara
; APPLICANT: Kao, Hung-Teh
; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 600-1-202 CIP
; CURRENT APPLICATION NUMBER: US/09/129,668B
; CURRENT FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 08/906,865
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 4
; LENGTH: 696
```



```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-129-668-4

```

Alignment Scores:

Pred. No.:	0.148	Length:	596
Score:	93.00	Matches:	45
Percent Similarity:	36.97%	Conservative:	167
Best Local Similarity:	27.27%	Mismatches:	64
Query Match:	8.72%	Indels:	41
DB:	4	Gaps:	7

US-09-920-953-2 (1-598) x US-09-129-668-4 (1-696)

Qy	81	GGCGGGCGAAGAGCATGAAGCTGCGGGTTGCACACCTTCTACGATAAGCTGCTGGCTGA	140
Db	530	GlyArgGlnSerArgProValAlaGlyGly-	539
Qy	141	CCGGGAGCTGCTCCCTTCTTCGAGTCCCTGGACATGCAAGACGACGAAGATGAAGCAGCT	200
Db	540	ProGlyAlaProProAlaAlaArgProProAlaSerProSerProGlnArgGlnAlaGly	559
Qy	201	CAAGTTTCATGAGCTTCGTTCGCGGACGACACCAATACAGGGCCGCAAGCATGTACCA	260
Db	560	ProPro-----GlnAlaThrArgGlnThrSerValSerGlyProAlaProProLys	576
Qy	261	CGACAC---CGCCCATCTGCTCAAGGGCCACGCGCTGGACCCACCGCCACTTTTGACAAGAT	317
Db	577	AlaSerGlyAlaProProGlyGlyGlnGlnArgGlnGlyProProGlnLysProProGly	596
Qy	318	CAAGCAGTACTCTGGAGAGACGCTGCAAGAGATGGG-----	353
Db	597	ProAlaGlyProThrArgGlnAlaSerGlnAlaGlyProValProArgThrGlyProPro	616
Qy	354	-----CGTCAAGCAGAGATGTGATCCAGCACG-----CGCGCGGAGTGG	391
Db	617	ThrThrGlnGlnProArgProSerGlyProGlyProAlaGlyAlaProLysProGlnLeu	636
Qy	392	TGGAGTCCACCCGCGAGAAATTTGACTTNCCTCAACAACTGCG-----	433
Db	637	AlaGlnLysProSerGlnAspValProProProAlaThrAlaAlaAlaGlyGlyProPro	656
Qy	434	CACCCAACCTGATTTTCAT-----TAACCCAAACCCAGCTCGAGCGCTCAT	478
Db	657	HisProGlnLeuPheAsnLeuProGluProAlaProProArgProSerLeuSerGlnAsp	676
Qy	479	TCCATCGATTTTGAG	493
Db	677	GluValLysAlaGlu	681

RESULT 4
5198348-1

```

; Patent No. 5198348
;
; APPLICANT: Bitter, Grant A.
;
; TITLE OF INVENTION: EXPRESSION OF EXOGENOUS POLYPEPTIDES
;
; AND POLYPEPTIDE PRODUCTS INCLUDING HEPATITIS B SURFACE
;
; ANTIGEN IN YEAST CELLS
;
; NUMBER OF SEQUENCES: 5

```

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/08/586,819
3 FILING DATE: 24-SEP-1990
4
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: 231,599
7 FILING DATE: 08-AUG-1988
8 APPLICATION NUMBER: 748,712
9 FILING DATE: 26-JUN-1985
10 APPLICATION NUMBER: 412,707
11 FILING DATE: 30-AUG-1983
12

```

```

; FILING DATE
; SEQ ID NO:1
; LENGTH: 226
5198348-1

```

Alignment Scores:

Pred. No.:	0.11	Length:	226
Score:	32.50	Matches:	43
Percent Similarity:	31.7%	Conservative:	18
Best Local Similarity:	22.40%	Mismatches:	53
Query Match:	8.24%	Indels:	78
DR:	6	Gaps:	11

US-09-920-953-2 (1-598) x 5198348-1 (1-226)

Qy	512	CGGCAACTGGCGCTCCCGCTCAAAATCGATGAATACGGCTCAGGCTGGGGTTCGGTT	451
		:	
Db	24	ArgileLeuThrIlePro-----GlnSerLeuAspSerTrp---TrpThr	37
Qy	452	AATGAAATCAGTTGGTGGCGCAG-----TTGTGGGNAAGTCAAATTCGTCGCGGGTG	399
		::: :	
Db	38	SerLeuAsnPheLeuGlyGlySerProValCysLeuGlyGlnAsnSerGlnSerProThr	57
Qy	398	GACTCCACCACTCCGCGGGGTGTGATCACATCTCTGTGACGCCCATCTCTTCGCAGC	339
		::: : :	
Db	58	SerAsnHisSerProThrSerCys-----	65
Qy	338	GTCCTCCAAGGTACTGCTTGTATCTTCTCAAGTGGCGGTGTCCAGGCGGTGGCCCTTG	279
Db	66	-----ProProlIleCys-----	71
Qy	278	ACCAGATGGGGTGTGCGTGTACATGCTTCGGCCCTTGTATTGGTGTGTCGCGCCAAAC	219
		: : : :	
Db	72	TyrArgTrpMetCysLeuArgArgPheIleIlePheLeuPhe-----	85
Qy	218	ACGAAGTCATGAATCTGACTGCTTCATCTTC-----	186
		:::	
Db	86	-----IleLeuLeuLeuCysLeuIlePheLeuValLeuLeuAspTyrGlnGly	102
Qy	185	-----TCTCTTTCATGCTCAGGACTCGAAGAGGCGACAGCTCCCGGTCA	138
		: : : : :	
Db	103	MetLeuProValCysProLeuIle-ProGlySerThrThrSerThrGlyPro-----	120
Qy	137	GCCAGCACCTTATCGTAGAGGTGTCAACGCCAGCT-----TC	99
		:	
Db	121	-----CysLysThrCysThrThrProAlaGlnGlyAsnSerMetPheProSe	136
Qy	98	ATGCGCTTCTGCGCGCCGAGCTCATCAACAGCTTCTTGGCGCCGATCCCGTTCGCGC	39
		: :	
Db	136	rCysCysCysThrLysProThrHisGlyAsnCysThrCys-----IleProlIleProSe	154
Qy	38	GTCT-----CTGTGGTGTCTGG	23
Db	154	rSerTirAlaPheAlaLysTyrLeuTrpGluTrp	165

RESULT 5 -

US-08-690-473-2
; Sequence 2, Application US/08690473

```

; Patent No. 5876923
;
; GENERAL INFORMATION:
;
; APPLICANT: Leopardi, Rosario
; APPLICANT: Rolman, Bernard
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICP4 AS AN
; TITLE OF INVENTION: INHIBITOR OF APOPTOSIS
; NUMBER OF SEQUENCES: 2
;

```

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P.O. Box 4433

CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

;
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/690,473
;
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   COMMENTS:

```

```

; FILING DATE: 26-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARCD:239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1298 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: linear
US-08-690-473-2

Alignment Scores:
Pred. No.: 0.212 Length: 1298
Score: 92.50 Matches: 51
Percent Similarity: 36.59% Conservative: 9
Best Local Similarity: 31.10% Mismatches: 48
Query Match: 8.67% Indels: 56
DB: 2 Gaps: 9

US-09-920-953-2 (1-598) x US-08-690-473-2 (1-1298)
QY 8 CTGCAGACGGCCACACACAGACGCGCGGAGCGGATCGGGCGCAAGAAGC 67
||| ||||| ||||| ||||| |||||
Db 156 LeuserProArgProAlaGlnProProArgArgArgHisGly----- 171
QY 68 TGTTCATGACCTGGCGCGCAGAGGATGAGCTGGCTGGTGTGACACCTTCTACGATA 127
||||| ||||| ||||| ||||| |||||
Db 172 -----ArgTTPArgProSerAlaSerSer----- 179
QY 128 AGGTGCTGCTGACCGGAGCTGCTCCCTTCTTCAGTCCCTGGACATGCAAGACAGA 187
||| ||| ||||| ||||| |||||
Db 180 -----ThrSerSerAspSerGlySerSer-----SerSerSer 191
QY 188 AGATGACAGCGTCAAGTTCATGACCTTCCTGTCGGCGGACACACCAATACAAGGCC 247
||| ||| ||||| ||||| |||||
Db 192 AlaSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 211
QY 248 GAAGCATGTACGACGACACACGCGCCATCTGTCACGAGGCGCCACG----- 290
||| ||| ||||| ||||| |||||
Db 212 Asp-HisAlaArgGluAlaArgAlaValGlyArgGlyProSerSerAlaAlaProAla 231
QY 291 -----CCTGGACACCGCCACTTTGACAAAGATCAAGCAGTACC 328
||||| ||||| ||||| ||||| |||||
Db 231 aProGlyArgThrProProProGlyProProProLeu-----SerGluAlaAlaPr 249
QY 329 TTGG---AGAGACGCTGCAAGAGAT-----GGCGCT----- 356
||| ||| ||||| ||||| |||||
Db 249 oLysProArgAlaAlaAlaArgThrProAlaAlaSerAlaGlyArgGluArgArgAr 269
QY 357 -----CAAGCAGGATGTATCCACGACGCGCGGAGTGTGGTGGTCCACCC 403
||||| ||||| ||||| ||||| |||||
Db 269 gAlaArgAlaAlaValAlaGlyArgAspAlaThrGlyArgPheThrAlaGly---GlnPr 288
QY 404 GCGACGAATT 413
||||| ||||| ||||| ||||| |||||
Db 288 oArgArgVal 291

RESULT 6
US-09-259-821A-2
; Sequence 2, Application US/09259821A
; Patent No. 6210926
; GENERAL INFORMATION:
; APPLICANT: LEOPARDI, ROSARIO
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICP4 IS AN INHIBITOR OF APOPTOSIS
; FILE REFERENCE: ARCD:317
; CURRENT APPLICATION NUMBER: US/09/259,821A

```

```

; CURRENT FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 08/690,473
; PRIOR FILING DATE: 1996-07-26
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1298
; TYPE: PRT
; ORGANISM: HERPES VIRUS, TYPE 1
US-09-259-821A-2

Alignment Scores:
Pred. No.: 0.212 Length: 1298
Score: 92.50 Matches: 51
Percent Similarity: 36.59% Conservative: 9
Best Local Similarity: 31.10% Mismatches: 48
Query Match: 8.67% Indels: 56
DB: 4 Gaps: 9

US-09-920-953-2 (1-598) x US-09-259-821A-2 (1-1298)
QY 8 CTGCAGACGGCCACACACAGACGCGCGGAGCGGATCGGGCGCAAGAAGC 67
||| ||||| ||||| ||||| |||||
Db 156 LeuserProArgProAlaGlnProProArgArgArgHisGly----- 171
QY 68 TGTTCATGACCTGGCGCGCAGAGGATGAGCTGGCTGGTGTGACACCTTCTACGATA 127
||||| ||||| ||||| ||||| |||||
Db 172 -----ArgTTPArgProSerAlaSerSer----- 179
QY 128 AGGTGCTGCTGACCGGAGCTGCTCCCTTCTTCAGTCCCTGGACATGCAAGACAGA 187
||| ||| ||||| ||||| |||||
Db 180 -----ThrSerSerAspSerGlySerSer-----SerSerSer 191
QY 188 AGATGACAGCGTCAAGTTCATGACCTTCCTGTCGGCGGACACACCAATACAAGGCC 247
||| ||| ||||| ||||| |||||
Db 192 AlaSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 211
QY 248 GAAGCATGTACGACGACACACGCGCCATCTGTCACGAGGCGCCACG----- 290
||| ||| ||||| ||||| |||||
Db 212 Asp-HisAlaArgGluAlaArgAlaValGlyArgGlyProSerSerAlaAlaProAla 231
QY 291 -----CCTGGACACCGCCACTTTGACAAAGATCAAGCAGTACC 328
||||| ||||| ||||| ||||| |||||
Db 231 aProGlyArgThrProProProGlyProProProLeu-----SerGluAlaAlaPr 249
QY 329 TTGG---AGAGACGCTGCAAGAGAT-----GGCGCT----- 356
||| ||| ||||| ||||| |||||
Db 249 oLysProArgAlaAlaAlaArgThrProAlaAlaSerAlaGlyArgGluArgArgAr 269
QY 357 -----CAAGCAGGATGTATCCACGACGCGCGGAGTGTGGTGGTCCACCC 403
||||| ||||| ||||| ||||| |||||
Db 269 gAlaArgAlaAlaValAlaGlyArgAspAlaThrGlyArgPheThrAlaGly---GlnPr 288
QY 404 GCGACGAATT 413
||||| ||||| ||||| ||||| |||||
Db 288 oArgArgVal 291

RESULT 7
US-08-843-659-2
; Sequence 2, Application US/08843659
; Patent No. 6218103
; GENERAL INFORMATION:
; APPLICANT: Leopardi, Rosario
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS US3 AND ICP4 AS
; NUMBER OF INVENTION: INHIBITORS OF APOPTOSIS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States

```

ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/843.659
APPLICATION NUMBER: US/08/843.659
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: ARSB:519
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1298 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-843-659-2

Alignment Scores:
Pred. No.: 0.212 Length: 1298
Score: 92.50 Matches: 51
Percent Similarity: 36.59% Conservative: 9
Best Local Similarity: 31.10% Mismatches: 48
Query Match: 8.67% Indels: 56
DB: 4 Gaps: 9

US-09-920-953-2 (1-598) x US-08-843-659-2 (1-1298)

```
Qy 8 CTGCAGCGCGCCACAGCAGCAGCGCGGAGCGGATGCGCGCGCAGAGAGC 67
Db 156 LeuSerProArgProGlnProProArgArgArgHisGly----- 171
Qy 68 TCTTTGATGACCTGGCGCGCAGAGCATGACCTGGCGGTTGACACCTCTTACGATA 127
Db 172 -----ArgTPrArgProSerAlaSerSer--- 179
Qy 128 AGGTCTGGTGCAGCGGAGCTGCTGCCCTTCTTCGAGTCCCTGACATGCAAGACGAGA 187
Db 180 -----ThrSerSerAspSerGlySerSer-----SerSerSer 191
Qy 188 AGATCAAGCAGTCAAGTTCATGACCTTCGTTGGCGGAGCAGACCAATACAGGCC 247
> 192 AlaSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 211
Qy 248 GAAGCATGTACGACGACACGCCCATCTGGTCAAGGGGCCACGG----- 290
Db 212 Asp-HisAlaArgGluAlaArgAlaValGlyArgGlyProSerSerAlaAlaProAlaAl 231
Qy 291 -----CCTGGACCCCGCCCATCTTGCAGCAAGATCAAGCAGTACC 328
Db 231 aProGlyArgThrProProProProGlyProProProLeu-----SerGluAlaAlaPr 249
Qy 329 TTGG---AGAGACGCTGCAAGAGAT-----GGCGCT----- 356
Db 249 oLysProArgAlaAlaAlaAlaArgThrProAlaAlaSerAlaGlyArgIleGluArgArg 269
Qy 357 -----CAAGCAGGATGTGATCCAGCAGCGCGCGGAGTGGTGGAGTCCAGCC 403
Db 269 gAlaArgAlaAlaValAlaGlyArgAlaThrGlyArgPheThrAlaGly---GlnPr 288
Qy 404 GCGACGAATT 413
Db 288 oArgArgVal 291
```

RESULT 8

US-09-413-814-79

Sequence 79, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Bayer, Stefan
APPLICANT: Bloeker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hofle, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413.814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 79
LENGTH: 1213
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-79

Alignment Scores:
Pred. No.: 0.233 Length: 1213
Score: 92.00 Matches: 54
Percent Similarity: 35.54% Conservative: 5
Best Local Similarity: 32.53% Mismatches: 63
Query Match: 8.62% Indels: 44
DB: 4 Gaps: 11

US-09-920-953-2 (1-598) x US-09-413-814-79 (1-1213)

```
Qy 12 AGACGCGGC-----CACCAG-----CACCACAGACGCGCGAAGC 47
Db 151 ArgArgGlyGluHisGlyProArgHisGlnProProLeuArgArgArgProGlyGlyPro 170
Qy 48 GGGATGCGCGCGCAAGAAGCTGTTGTGATGACCT-----GGG 83
Db 171 GlyAlaArgAlaLeuValAlaGluLeuArgProValGlyLeuArgArgValArgAspAla 190
Qy 84 CGGCGCAGAGCATGAAGCTGGCGGTTGACACCTTCTACGATAAGGTGCTGCTGACCC 143
Db 191 ArgArgArgArgArgArgArgArgProArgProTyrProArgLeuGlySerGly----- 208
Qy 144 GGAGCTGCTGCCCTTCTTTCGAGTCCCTGGACATGCAAGACAGAGAAGTAAAGCA----- 197
Db 209 -----AlaLeuAlaArgAlaArgGly---AlaArgAlaGlyAspArgValGluLeu 224
Qy 198 -----GGTCAAGTTCATGAGTCTGCTGTTGGCGGAGCAGACCA----- 236
Db 225 GlyProGlyAlaAspGlyAspAlaHisGlyArgValProArgArgGlyArgProGlyAla 244
Qy 237 ---ATACAGGCGCGCAAGCATGTACGACGACACACCCCATCTGTGTCAGGCGCAGCGCT 293
Db 245 ValValAlaProProArgHisAspGluArgArgLeuAspProAlaGluAlaProArgSer 264
Qy 294 GCACCACCCCATTTTGACAAGATCAAGCAGTACCTTGGAGAGAGAGCTGCCAAGAGATCGG 353
Db 265 HisProArgGlyLeu-----ProArgAlaProArgArgGluProArgArg----- 279
Qy 354 CGTCAGCAGGATGTGAT-----CCAGCAGCGCGCGGAGTGGTGGAGTCC 398
Db 280 ArgAspArgGlyValAspLeuValAspArgProProAspArgGlyArgArgProGlyVal 299
Qy 399 -----CACCGCGCAGCA 410
```


APPLICANT: Chou, Joany
TITLE OF INVENTION: Method for Treating Tumorigenic Diseases
TITLE OF INVENTION: Diseases
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago

Qy	280	AAGGCCACGCGCTGGACACCGCCGCCCTTTGACAACGATCAACGACGAGTACCTTGGAGACGACG	339
Db	252	iProGluAlaAaIaTrpValPheArgGlyAlaLapProGlySerSerAlaProSerArgSerPr	272
Qy	340	CTGCAAGAGATGGCGCTCAACGACGAGATGTCATCC-----ACGACCGCCCGCGAGTGGTG	393
Db	272	oGluuArgTrpGlnGluProArgIleTyThrLeuGlyAlaSerProProSerGlnGlnI	292
Qy	394	GAGTCCACCC	403
Db	292	vGlyProPro	295

RESULT 12

US-09-283-471A-41
; Sequence 41, Application US/09283471A
; Patent No. 6340673
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; APPLICANT: Chou, Joany
; TITLE OF INVENTION: Method For Treating Tumorigenic Diseases
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/283,471A
; FILING DATE: 04-APR-1999

US-09-920-953-2 (1-598) x US-09-283-471A-41 (1-355)

```
QY 17 CGGCCACACACAGACGCGGCGGAGCGGGTTCGGGCGCCCAAGAGC----- 67
Db 165 ArgProProArgProProArgProProArgProProArgProProArgProProArgProProThr 184
QY 68 -----TGTTTATGACCTGGCGCGGCGGAGGCGGATGAAGCTGGCGGTTCACACCTTCT 121
Db 185 SerGlyCysAlaThrTrpSerGlyProProArgProProAlaTrpArgAlaAlaArg 204
QY 122 ACATAAGGTGTGGCTGACCGGAGCTGCTGCCCTTCCTTCAGTCCCTGGACATGCAAG 181
Db 205 Gly-----ProAlaSerGlyProThrGlyLeuGly 214
QY 182 AGC-----AGAAGATGAACGAGGTCAGTTCATGAGTTCCTGTTGGCGGAG 229
Db 215 SerGlyAlaGlyTrpArgProProArgProProArgProProArgProProArgProProThr 184
QY 230 CAGACCAATACAGGCGCG-----AAGCATGTACGACGACGACGACGACGACGACGACG 232
Db 232 oArgProValProGlyProProGluProProAlaGluProAlaGluProAlaGluProAlaGluPro 252
QY 394 GAGTCCACCC 403
Db 292 yGlyProPro 295

RESULT 13
PCT-US91-06532-3
; Sequence 3, Application PC/TUS9106532
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; TITLE OF INVENTION: Recombinant Herpes Simplex Viruses
; TITLE OF INVENTION: Vaccines and Methods
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESS: Bicknell
; STREET: Two First National Plaza Suite 2100
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/06532
; FILING DATE: 19910910
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gruber, Lewis S.
; REGISTRATION NUMBER: 30,060
; REFERENCE/DOCKET NUMBER: 27373/8235
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/346-5750
; TELEFAX: 312/984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US91-06532-3

Alignment Scores:
Pred. No.: 0.307 Length: 355
Score: 89.00 Matches: 41
Percent Similarity: 38.89% Conservative: 15
Best Local Similarity: 28.47% Mismatches: 61
Query Match: 8.34% Indels: 28
DB: 5 Gaps: 6

US-09-920-953-2 (1-598) x PCT-US91-06532-3 (1-355)
QY 17 CGGCCACACACAGACGCGGCGGAGCGGGTTCGGGCGCCCAAGAGC----- 67
Db 165 ArgProProArgProProArgProProArgProProArgProProArgProProArgProProThr 184
QY 68 -----TGTTTATGACCTGGCGCGGCGGAGGCGGATGAAGCTGGCGGTTCACACCTTCT 121
Db 185 SerGlyCysAlaThrTrpSerGlyProProArgProProAlaTrpArgAlaAlaArg 204
QY 122 ACATAAGGTGTGGCTGACCGGAGCTGCTGCCCTTCCTTCAGTCCCTGGACATGCAAG 181
Db 205 Gly-----ProAlaSerGlyProThrGlyLeuGly 214
QY 182 AGC-----AGAAGATGAACGAGGTCAGTTCATGAGTTCCTGTTGGCGGAG 229
Db 215 SerGlyAlaGlyTrpArgProProArgProProArgProProArgProProArgProProThr 184
QY 230 CAGACCAATACAGGCGCG-----AAGCATGTACGACGACGACGACGACGACGACGACG 232
Db 232 oArgProValProGlyProProGluProProAlaGluProAlaGluProAlaGluProAlaGluPro 252
QY 394 GAGTCCACCC 403
Db 292 yGlyProPro 295

US-09-413-814-78
; Sequence 78, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-413-814-78
```

Alignment Scores:

Pred. No.: 0.489 Length: 882
 Score: 88.50 Matches: 68
 Percent Similarity: 33.73% Conservative: 16
 Best Local Similarity: 27.31% Mismatches: 63
 Query Match: 8.29% Indels: 102
 DB: 4 Gaps: 14

US-09-920-953-2 (1-598) x US-09-413-814-78 (1-882)

```

QY 20 CCACACACACACAG-----AGACGGCGGAAGCGGATCGCGCGCAAGAGCTGTTG 73
Db 590 ProGluAlaProArgAlaArgHisArgAlaArgAlaProArgValArgArg 609
QY 74 ATGACCTGGCGGCGCC---ACAAGCGATGAAGCTGGCGGTGCACCTTCTACGATAGG 130
Db 610 LeuVal-GlyArgArgLeuArgArgAlaArgAlaLeuArgArgLeuArgAlaG 629
QY 131 T-----GCTGCTGACCGCGAGCT----- 149
b 629 yProAlaPheProAlaAlaGlyAlaProGlyAlaValArgAlaArgArgSerProAl 649
QY 150 ----GCTGCCCTTCTCCGATCCCTGGACATGCAAGAGCA-----GAAGATGAAGC 196
Db 649 aGlyValAlaValArgArgGlyProGly---GlyArgAlaProValLeuAspGluAl 668
QY 197 A----- 197
Db 668 aLeuGlyArgAlaAlaProGlyProArgAlaGlyArgProArgAlaGlyGlyAlaLysAs 688
QY 198 -----GGTCAAGTTCATGAG----- 212
Db 688 pValAlaArgGlyAlaAlaGluProArgArgArgGlyArgAlaHisProProAspGlnG 708
QY 213 ----CTTCGTGTTTGGCGGAGCAGACCAATACAAGGCCGCGGATGTACGACGACAGC 268
Db 708 yValLeuArgAlaGlyGlyAspLeuProLeuHisGlyProAlaArgArgValGlnAlaAr 728
QY 269 CCCATCT-----GGTCAAGGGCCAGCGCCCTGGA-----CCACC 301
Db 728 gProAlaProAlaHisArgAlaArgGlyProArgArgGlyHisArgArgGluProG 748
QY 302 GCCA-----CTTTGACAGATCAAGCAGTACCTTGAG 334
Db 748 uProArgArgAspArgAlaAspArgLeuLeuArgGlnProAlaArgAlaProAspAr 768
QY 335 ACACCGCTGCAAGA---GATGGCGCTCAAGCAGGATGTATCCAGCAGCGCGCGAGTGG 391
Db 768 gLeuArgGlyArgProAspValArgArgAlaGlyProProArgAlaArgArgAlaAr 788
QY 392 TGGAGTCCACCGCGCAGCAATTTGACTTNTCCCAACAACACTGCGCACCACTGATTTTCAT 451
Db 788 gGlyValArg----- 791
QY 452 TAACCCAAACCCAGCCTGAGCGCTCATTTCCATCGATTTTGGCGGGAGCGCCAGTTGCC 511
Db 792 -----AlaProGlyProAlaValArgProAla-----ArgArgGlyPr 804
QY 512 GAGCGCGCCAGGGGCGCCAGGAGC 536
Db 804 oAlaAlaGluGlyGlyArgArgAla 812

```

RESULT 15
 PCT-US96-10602-6

; Sequence 6, Application PC/TUS9610602

; GENERAL INFORMATION:

; APPLICANT: The General Hospital Corporation

; TITLE OF INVENTION: INHIBITION OF HEPATITIS B REPLICATION

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM: disk
 MEDIUM TYPE: Floppy
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/10602
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/017,814
 FILING DATE: 20-JUN-1995
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Clark, Paul T.
 REGISTRATION NUMBER: 30,162
 REFERENCE/DOCKET NUMBER: 00786/282001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 397 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US96-10602-6

Alignment Scores:
 Pred. No.: 0.409 Length: 397
 Score: 88.00 Matches: 49
 Percent Similarity: 30.90% Conservative: 23
 Best Local Similarity: 21.03% Mismatches: 62
 Query Match: 7.84% Indels: 99
 DB: 5 Gaps: 13

US-09-920-953-2 (1-598) x PCT-US96-10602-6 (1-397)

```

QY 563 CAGCAAGCGGTGGCAACAGATTTCAGCGCTCCTGGCGCCCTGGCGCGCTCGCAACTG 504
Db 151 ArgArgGlyArgSerProArgArgArgThrProSerPro-----ArgArgArgArgSer 168
QY 503 CGCGTCCCGCTCAAAATCGATGGAATCAGCGCTCAGCG----- 465
Db 169 GlnSerProArgArgArgArgSerGlnSerArgLeuGlyProLeuLeuValLeuGlnAla 188
QY 464 -----TGGGTTGG 456
Db 189 GlyPhePheLeuLeuThrArgIleLeuThrIleProGlnSerLeuAspSerTrp---Trp 207
QY 455 GTTAATGAAATCAATGTTGGTGGCGCAG-----TTGTTGGGNAAGTCAAAATTCGTCGCGG 402
Db 208 ThrSerLeuAsnPheLeuGlyGlyThrValCysLeuGlyGlnAsnSerGlnSerPro 227
QY 401 GTGGACTCCACCACTCCGCGGCGGTGCTGGATCATCATCTCTTTCAGCCCATCTCTTTC 342
Db 228 ThrSerAsnHisSerProThrSerCys-----ProProThrCys 240
QY 341 AGCTCTCTCTCAAGGTACTGCTTTCATCTTGTCAAGTGGCGGTGCTCCAGCCGCTGGCCCC 282
Db 241 -----ProGlyTyr----- 243
QY 281 TTGACCAAGATGGCGGTGTGCGTGTACATGCTTGGCGCTTGTATTGTGCTGCTCCGCCA 222
Db 244 -----ArgTrpMetCysLeuArgArgPheIlePheLeuPhe----- 256
QY 221 AACACGAAGCTCATGAACCTTGACCTGCTTTCATCTTC----- 186
Db 257 -----IleLeuLeuLeuCysLeuIlePheLeuValLeuLeuAspTyrGln 272

```

```

QY 185 -----TGCTCTTCATGTCAGGACTCGAAGAGGCGCAGCAGCTCCGGG 141
Db 273 GlyMetLeuProValCysProLeuIle-ProGlySerSerThrThrSerThrGlyPro-- 291
QY 140 TCAGCCAGCACCTTATCGTAGAGGTGTCAACGCCAGCT----- 101
Db 292 -----CysArgThrCysMetThrThrAlaGlnGlyThrSerMetTyrPr 306
QY 100 -TCATGCCCTTCTGCGCGCCAGGTCAACACAGCTTCTTGGCCCGCATCCCGTTCC 42
Db 306 oSerCysCysCysThrLysProSerAspGlyAsnCysThrCys-----IleProIlePr 324
QY 41 GCCGTCT-----CTGTGGTCTGG 23
Db 324 oSerSerTrpAlaPheGlyLysPheLeuTrpGluTrp 336

```

Search completed: April 16, 2003, 13:02:35
 %b time : 20.5 secs


```

US-09-920-953-2 (1-598) x US-09-821-877-2 (1-389)
OY 581 GCTGATGCTCTTCAATACAGCAAGGCTGGCAACGATTTGACGGCTCTCGGGCCCT 522
Db 132 AlaGlySerSer-----SetGlyThrValAsnProValProThrThrValSerPro 149
OY 521 GGGCCCGCT-----CGCAACTGGCGCTCCCGCTCAAAATCGAATGAGCGTCA 468
Db 150 IleSerSerIlePheSerArgIleGlyAspProAlaArgAsnMetGluAsnIleThrSer 169
OY 467 GGC-----465
Db 170 GlyPheLeuGlyProLeuLeuValLeuGlnAlaGlyPhePheLeuLeuThrArgIleLeu 189
OY 464 -----TGGGGTTGGGTTAATGAAATACAGTTGGTGGCGCGCAG 429
Db 190 ThrIleProGlnSerLeuAspSerTrp---TrpThrSerLeuAsnPheLeuGlyGlyThr 208
OY 428 -----TTGTTGGGNAAGTCAAAATTCGCGCGGTGGACTCCACCACTCCCGCGCGTGC 375
Db 209 ThrValCysLeuGlyGlnAsnSerGlnSerProThrSerAsnHisSerProThrSerCys 228
OY 374 TGGATCACATCTCTTGCAGCCCATCTCTTGCAGCGTCTCTCCAAAGTACTGCTTGATC 315
Db 229 -----ProProThrCys-----ProGlyTyr-----235
OY 314 TTGTCAAAGTGGCGGTGGTCCAGCGCGCTTGACGAGTGGCGGTGTGCGTCGTAC 255
Db 236 -----ArgTrpMetCysLeuArgArg 242
OY 254 ATGCTTCGGCCCTTGATTTGCTGCTGCCGCCAAACGAGCTCATGACCTGCG 195
Db 243 PheIleIlePheLeuPhe-----IleLeuLeuLeuCys 253
OY 194 TTCATCTC-----TGTCTTGTGATG 174
Db 254 LeuIlePheLeuLeuValLeuLeuAspTyrGlnGlyMetLeuProValCysProLeuIle 273
OY 173 TCCAGGACTCGAAGAGGCGAGCGTCCGGGTGACGACGACCTTATCTAGAGGTG 114
Db 274 -ProGlySerSerThrThrSerThrGlyPro-----CysArgAlaCys 287
OY 113 TCAACCCCGCAGCT-----TCATGCTTCTGCGCGCCCGCCAGGTCA 75
Db 287 sThrThrProAlaGlnGlyThrSerMetTyrProSerCysCysThrLysProSerAs 307
OY 74 TCAACAGCTTCTTGGCGCGCGCATCCGCTTCCGCGCTCT-----CT 33
Db 307 pGlyAsnCysThrCys-----IleProIleProSerSerTrpAlaPheGlyLysPheLe 325
OY 32 GTGCTGCTGG 23
Db 325 utrpGluTrp 328

RESULT 2
US-09-925-288A-2
; Sequence 2, Application US/09825288A
; Publication No. US20020192822A1
; GENERAL INFORMATION:
; APPLICANT: LEOPARDI, ROSARIO
; APPLICANT: ROIZMAN, BERNARD
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICP4 IS AN INHIBITOR OF APOPTOSIS
; FILE REFERENCE: ARCD:317USC1
; CURRENT FILING DATE: 2001-04-02
; PRIOR FILING DATE: 2001-09-25, 821
; PRIOR FILING DATE: 1999-03-01
; PRIOR FILING DATE: 1996-07-26
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2

```

```

; LENGTH: 1298
; TYPE: PRT
; ORGANISM: HERPES VIRUS, TYPE 1
US-09-825-288A-2
Alignment Scores:
Pred. No.: 1.97
Score: 92.50
Percent Similarity: 36.59%
Best Local Similarity: 31.10%
Query Match: 8.67%
Gaps: 9
Length: 1298
Matches: 51
Conservative: 9
Mismatch: 48
Indels: 56
Gaps: 9
US-09-920-953-2 (1-598) x US-09-825-288A-2 (1-1298)
OY 8 CTGACAGCGCGGCCACACAGAGCGGCGGAGCGGATCGGGCGCAAGRGC 67
Db 156 LeuSerProArgProProAlaGlnProProArgArgArgArgHisGly-----171
OY 68 TGTGTGATGACCTCGCGCGCGCAGAGGCATGAAGCTGGCGTTTCACACCTTCTACGATA 127
Db 172 -----ArgTrpArgProSerAlaSerSer---179
OY 128 AGGTGCTGCTACCGCGAGCTGCTGCCCTTCTTCGAGTCCCTCGACATGCAAGACGAGA 187
Db 180 -----ThrSerSerAspSerGlySerSer-----SerSerSer 191
OY 188 AGATGAAGCAGGTCAAGTTGATGAGCTTCGTGTTGGCGGAGCAGACCAATACAGGGCC 247
Db 192 AlaSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 211
OY 248 GAACGCTACGACGCGACGCGCCATCTGCTCAAGGCGCCACGG-----290
Db 212 Asp-HisAlaArgGluAlaArgAlaValGlyArgGlyProSerSerAlaAlaProAlaAl 231
OY 291 -----CCTGCACACCCCTTTCACAAGATCAAGCAGTACGACGAC 328
Db 231 aProGlyArgThrProProProGlyProProLeu-----SerGluAlaAlaPr 249
OY 329 TTGG---AGAGACGCTCAAGAGAT-----GGGCGT-----356
Db 249 OlySerProArgAlaAlaAlaArgThrProAlaAlaSerAlaGlyArgIleGluArgArg 269
OY 357 -----CAAGCAGGATGTATCCAGCAGCGCGCGGAGTGGTGGAGTCCACCC 403
Db 269 GAlaArgAlaAlaValAlaGlyArgAspAlaThrGlyArgPheThrAlaGly---GlnPr 288
OY 404 GCGACGAAAT 413
Db 288 oArgArgVal 291

RESULT 3
US-09-903-456-77
; Sequence 77, Application US/09903456
; Patent No. US20020138874A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Mukerji, Pradip
; APPLICANT: Leonard, Amanda Eun-Yeong
; APPLICANT: Huang, Yung-Sheng
; APPLICANT: Pereira, Suzette L.
; TITLE OF INVENTION: ELONGASE GENES AND USES THEREOF
; FILE REFERENCE: 6407.US.P3
; CURRENT FILING DATE: 2001-07-11
; PRIOR FILING DATE: 2001-07-11
; PRIOR FILING DATE: 2000-07-24
; PRIOR FILING DATE: 1999-08-23
; PRIOR FILING DATE: 1998-09-02
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: FastSeq for Windows Version 4.0

```



```
Alignment Scores: 2.26 Length: 175
Pred. No.: 89.00 Matches: 50
Score: 36.36% Conservative: 22
Percent Similarity: 25.25% Mismatches: 74
Best Local Similarity: 7.93% Indels: 52
Query Match: 9 Gaps: 11
DB:

US-09-920-953-2 (1-598) x US-10-012-542-472 (1-175)
QY 533 CTTGGCCCGCTGGCGCTCGCAACTGGCGCTCCCGCTCAAAATCATGATGAG 474
Db 9 ProGlyLysProThrArg---ArgGlyLeuGlyLeuProThrAlaSer----- 23
QY 473 CGCTCAGCTGGGGTTGGTTAATGAAATACAGTTGGTGGCGCAGTTGTTGGGNAAGTCA 414
Db 24 ---SerGlyTrpValTrp----- 28
QY 413 AATTCTCGCGGCTGAGTCCACCCTCGCGCGCGCTGGATCATCATCTGCTGTGAGC 354
Db 29 ArgArgCysIleAlaSerTrpGlyThrAlaThrAlaAlaTrpProCysSerCysGlyThr 48
QY 353 CCATC-----TCTTGCAGCGTCTCTCCAAGTACTGCTGTGATCTGTCAAGTGG 303
Db 49 GlyMetAlaThrProSerCysSerPro-----CysThrTrpValAlaArgThr 66
QY 302 CGGTGTCCAGGCGCTGGCCCTTGACAGATGGCGTGTGCTGTGATGCTTCCGGCC 243
Db 67 Arg-----Profile-----AlaCysSerSerLeuHisProTrpPro 78
QY 242 TWTATTGTCTGCTCCGCCCAACACGAGCTCATGAAC-----TTGACC 198
Db 79 AlaSerTrpAlaProProSerHisProAlaAlaSerProTyProSerProLeuGly 98
QY 197 TGTCTCATCTTCTGCTTTCATGTCAGGAGTCCAGGAGTCCGAGGCGGAGC---AGCTCCGG 141
Db 99 ThrArgIleThrSerAlaGlyThrArgThrAlaProArgAlaSerLeuGluAlaGly 118
QY 140 TCAGCCAGCACCTTATCGTAGAGGTGTCAACCGCAGCTTCTGCTGCGCGCC 81
Db 119 GlyLeuAlaProAlaAlaIleProThrPheAsnGlyProValLeuPro---AlaProSer 137
QY 80 AGCTCATCAACAGCTTCTTGCC----- 57
Db 138 HisSerSerGlyArgSerLeuArgGluSerSerGlyArgProAlaGlyArgTyTrp 157
QY 56 CCGCATCCCGCTTCCCGCTCTCTGTTGCTGTGGTGGCGCGCTCTCCAGGACG 3
Db 158 ProLeuGlnAlaThrThrMetLeuIleGlnProMetAlaAlaGluAlaAlaSer 175

RESULT 9
US-10-102-806-543
; Sequence 543, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 543
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens

FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (154)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (167)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-543
Alignment Scores: 3.54 Length: 352
Pred. No.: 88.00 Matches: 54
Score: 35.93% Conservative: 29
Best Local Similarity: 23.38% Mismatches: 80
Query Match: 9 Indels: 68
DB: 11

US-09-920-953-2 (1-598) x US-10-102-806-543 (1-352)
QY 14 AGCGGCCACCAAGCA-----CCACAGACGCGGGAAGGGGATCGGGCGCAAGA 64
Db 10 ThrArgProMetAlaAlaGluProGlnGlnGlnLysGlnGluProLeuGlySerAsp 29
QY 65 AGCTGTTGATGACC-----TGGCGCGCGCAGAGGCATGA 100
Db 30 SerGluValLeuThrValTrpProMetLysProSerTrpPleuSerArgThrGluPhe 49
QY 101 AGCTGGCGGTTGACACCTTCTACGATAAGTGTC-----TGGTACCGCGAGCTGC 151
Db 50 SerLysArgLeuLeuCysArgThrLeuTrpCysGlnSerGlyTrpSerSerArgSerTy 69
QY 152 TCCCTTCTTCAGTCCCTGGACATGCAAGACGACAGATGAAGCAGGTCAAGTTCATGA 211
Db 70 ThrArgSerMetLeuLysMetThrThrSerIleAsnArgSerArgThrSerThr--- 88
QY 212 GCTTCGTGTTGGCGGAGCAGCAATACAAGGCGGAGCATGTACGACGACACGCCCC 271
Db 89 -----LysSerThrArgThrSerAlaArgProGlyLeuThr----- 100
QY 272 ATCTGTCGAAGGCGCCGCTGGACCGCCCTTTCACAAAGATCAAGACGATACCTTG 331
Db 101 -----AlaThrValSerIleGlyLeuSerAspSerProThr 112
QY 332 GAGACACGCTGCAAGAGATGGCGTCAAGCAGGATGTATCCAGCACGCGCCGAGTGG 391
Db 113 TrpArgHisCys-----TrpMetThrAlaArgSerCysSerGlyCluLysGlyGlyHis 130
QY 392 TGGAGTCCACCGCGCAGCAATTG-----ACT 418
Db 131 TrpAlaProArgGlnValGlyValTrpLeuLeuProGlyArgValGlyCysValSerSer 150
QY 419 TNCACAACACTGCGCACCCCACTGATTTTCATTAAACCAACCCAGCCTGAGCCTCAT 478
Db 151 ArgValSer---SerPheProGlyAsp----- 159
QY 479 TCCATCGAT-----TTTGAGCGGGAGCGCCAGTTCGCGACGCGCCAGGGGCC 529
Db 160 GlyLeuAspSerGlyLeuAla---ArgGlySerAlaValSerAlaLeuAlaSerGlyLeu 179
QY 530 CAGGAG-----CCTGCAATCTGTTTCCAGCCCTTCTGCT 562
Db 180 ValGluGluProMetLeuGlyProProPheHisProThrProArgPheLysAlaValSer 199
QY 563 GCATTGAAGACCATCAGCCATTTTCGCCACCAA 595
Db 200 Ala---LysSerLysGluAspLeuValSerGln 209

RESULT 10
US-09-812-862-6
; Sequence 6, Application US/09812862
```


OY 434 CCCGAGTTGTTGGGNAAGTCAAATTCCTCGCGGGTGCACCTCCACCACTCGGGG---GCG 378
Db 1477 -----GlnGluArgSerArgThrHisLeuThrValProAlaGlySer 1490
OY 377 TGTGGATACATCCCTGCTTACGCGCCATCTCTTGACGCTCTCTCAAGGTACTGCTTG 318
Db 1491 ThrGlnLeuProLeuCysProGlyLeuPheProCysGlyValAlaProGlyLeuCysLeu 1510
OY 317 ATCT-----TGTCNAAGTGGC----- 302
Db 1511 -ThrProGluGlnLeuCysAspGlyIleProAspCysProGlnGlyGluAspGluLeuAs 1530
OY 301 -----GCTGTCAGCGCCGTCGCG-----CCCTGACCATGCG 270
Db 1530 pCysGlyGlyLeuProAlaLeuGlyGlyProAsnArgThrGlyLeuProCysProGlu-- 1549
OY 269 GCGTGTGCTGTCATATGC---TTCGGCCCTTGTATTGTCTGCTGCTCCGCAACACGAAG 213
Db 1550 -----TyrThrCysProAsnGlyThrCysIleGlyPhe----- 1560
OY 212 CTCATGAATTCACCTGCTTCATCTCTGCTCTTGTGCTTCCAGGACTCGAAGAGGCG 153
Db 1561 -----GlnLeuValCysAspGly---GlnProAspCys 1570
OY 152 AGCAGCTCCGGGTACG-----CCAGCACCTTATCGT----- 122
Db 1570 sGlyArgProGlyGlnValGlyProSerProGluGlnGlyCysGlyAlaTrpGlyPr 1590
OY 121 -----AG 120
Db 1590 oTrpSerProTrpGlyProCysSerArgThrCysGlyProTrpGlyGlnGlyArgSerAr 1610
OY 119 AGGTGTCAACCGCCAGCTTCATGCTCTGCGCCGC-----CCAGGTTCATCAAC 69
Db 1610 gArgCysSerProLeuGlyLeuLeuValLeuGlnAsnCysProGlyProGluHisGlnSe 1630
OY 68 ACTTCTTGGCGCCGATCCGCTTCCGCG-----TCTCTGTGCTGCTGCTG 21
Db 1630 rGlnAlaCysPheThrAlaAlaCysProValAspGlyGluTrpSerThrTrpSerProTr 1650
OY 20 G 20
Db 1650 p 1650

RESULT 12
US-09-956-425-4
; Sequence 4, Application US/09956425
; Patent No. US20020045192A1
GENERAL INFORMATION:
; APPLICANT: Kriwacki, Richard
; APPLICANT: Bothner, Brian
; APPLICANT: Lewis, William
; TITLE OF INVENTION: Arf and Hdm2 Interaction Domains and Method of Use Thereof
; FILE REFERENCE: 1340/1/035
; CURRENT APPLICATION NUMBER: US/09/956,425
; CURRENT FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-956-425-4

Alignment Scores:
Pred. No.: 2.89 Length: 132
Score: 87.50 Matches: 50
Percent Similarity: 35.48% Conservative: 16
Best Local Similarity: 26.88% Mismatches: 53
Query Match: 67
DB: 10 Gaps: 10

US-09-920-953-2 (1-598) x US-09-956-425-4 (1-132)

OY 595 TTGTTGCGAAATGGCTGATGGCTCTTAATGCAGCAAGGCTGGCAACGATTTCCAGG 536
Db 1 MetValArgArgPheLeuValThrLeuArg-----IleArgArg 13
OY 535 CTCCTGCGCCCTCGGCGCG-----CTCGGCAAC 506
Db 14 AlaCysGlyProProArgValArgValPheValValHisIleProArgLeuThrGlyGlu 33
OY 505 TGGCGCTCCCGCTCAAAATCGATGAGCGCTCAGCGCTGGGTGGTTAATGAA 446
Db 34 TrpAlaAlaPro-----GlyAlaProAlaAlaValAlaLeuVal--- 46
OY 445 ATCAGTTGGTGGCGCAGTTGTTGGGNAAGTCAAATTCCTCGCGGGTGGACTCCACCATC 386
Db 47 LeuMetLeuLeuArgSer-----GlnArgLeuGlyGlnGlnProLeu 60
OY 385 CGCGCGCGTCTGGATCACATCTGCTTGCAGCCCATCTCTTGCACGCTCTCTCCAAGCT 326
Db 61 ProArgArgProGlyHisAspAspGlyGlnArgProSerGlyGlyAlaAla--- 77
OY 325 ACTGCTTGATCTGTCAAAGTGGCGTGTCCAGGCGCGTGTGACCATGATGGCGT 266
Db 78 -----AlaAlaProArgArgGly----- 83
OY 265 GTGCGTGTACATGCTTCGCGCCCTTGTATTGCTCTGCTCCGCCCAACAGCAAGCTATGA 206
Db 84 -----AlaGlnLeuArgArgProArgHisSer--- 92
OY 205 ACTTGACCTGCTTCTCTGCTCTGCTGCTTCCAGG-----ACTCGAAGAGG 155
Db 93 -----HisProThrArgAlaArgArgCysProGlyGlyLeuProGlyHisAlaGly 109
OY 154 GCAGCAGCTCCGGTTCAGCGACCTTATCGTAGAGGTGTCAACCGCATCTTCATGC 95
Db 110 GlyAlaAlaProGlyArgGlyAlaAlaGlyArgAlaArgCysLeuGlyProSerAla--- 128
OY 94 CTTCTGCGCGCCCGCAGGT 77
Db 129 -----ArgGlyProGly 132

RESULT 13
US-09-220-920-116
; Sequence 116, Application US/09220920
; Patent No. US20020002269A1
GENERAL INFORMATION:
; APPLICANT: Milbrandt, Jeffrey D.
; APPLICANT: Baloh, Robert H.
; TITLE OF INVENTION: Artemin, A No. US20020002269A1e1 Neurotrophic Factor
; FILE REFERENCE: 6029-7996
; CURRENT APPLICATION NUMBER: US/09/220,920
; CURRENT FILING DATE: 1998-12-24
; EARLIER APPLICATION NUMBER: 09/163,283
; EARLIER FILING DATE: 1998-09-29
; EARLIER APPLICATION NUMBER: 60/108,148
; EARLIER FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/218,698
; EARLIER FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 116
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-220-920-116

Alignment Scores:
Pred. No.: 3.31 Length: 201
Score: 87.50 Matches: 55
Percent Similarity: 31.75% Conservative: 12
Best Local Similarity: 26.07% Mismatches: 93
Query Match: 7.80% Indels: 51
DB: 10 Gaps: 8


```

Db      24  AlAGlyClySerSer-----SerGlyThrValAsnProValProThrThrAlaSerPro 41
QY      521  GGGCGGCGCT-----CGGCAACTGGCGCTCCCGCTCAAAATCGATGGAATGAGCGCTCA 468
Db      42  IleSerSerIlePheSerArgThrGlyAspProAlaProAsnMetGluAsnThrThrSer 61
QY      467  GGC-----465
Db      62  GlyPheLeuGlyProLeuLeuValLeuGlnAlaGlyPhePheLeuLeuThrArgIleLeu 81
QY      464  -----TGGGGTTGGGTTAATCAAAATCAGTTGGGTGCG---432
Db      82  ThrIleProGlnSerLeuHisSerTrp---TrpThrSerLeuAsnPheLeuGlyAlaAla 100
QY      431  ---CAGTTGTGGNAAGTCAAAATTCGTGCGGGTGGACTCCACCACTCCGGCGGCGTGC 375
Db      101  ProThrCysLeuGlyGlnAsnSerGlnSerProThrSerAsnHisSerProThrSerCys 120
      374  TGGATCACATCTCTGTTGACGCCCATCTCTTGCAGCGTCTCTCCAAAGGTACTGCTTGATC 315
      121  -----ProIleCys-----124
QY      314  TTGTCAAAGTGGCGGTGGTCCAGGCGCTGGCCCTTCACCAGATGGCGGTGTCGTCGTAC 255
Db      125  -----ProGlyTyrArgTrpMetCysLeuArg 134
QY      254  ATGCTTCGGCCCTTGTATTGGTCTGCTCGGCCAAACACAGCAAGCTCATGAACCTTGACCTGC 195
Db      135  PheIleIlePheLeuPhe-----IleLeuLeuLeuCys 145
QY      194  TTCATCTTC-----TGCTCTTGCATG 174
Db      146  LeuIlePheLeuLeuValLeuLeuAspTyrGlnGlyMetLeuSerValCysProLeuLeu 165
QY      173  TCCAGGACTCGAAGAGGCGACAGCTCCGGGTGTCAGCCAGCACCTTATCGTAGAAGGTG 114
Db      166  ProArgThrSerThr-----ThrSerThrGlyProCysLysThrCysThrIleProAla 183
QY      113  TCAACCGCCAGCTTCATGCTTCTGCG---CGGCCAGGTCATCAACAGCTTCTTGGCG 57
Db      184  GlnGlyThrSerMetPheProSerSerCysCysThrLysProSerAsp-GlyAsnCysTh 203
QY      56  CCGCATCCGCTTCCCGCGTCT-----CTGTGGTGCTGG 23
Db      203  rCysIleProIleProSerSerTrpAlaPheAlaArgPheLeuTrpGluTrp 220

```

arch completed: April 16, 2003, 13:03:25
 job time : 29 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2003, 11:24:35 ; Search time 2154 Seconds
(without alignments)
8079.609 Million cell updates/sec

Title: US-09-920-953-2
Perfect score: 598
Sequence: 1 ggcgtcctgcagacgccc.....agccatttcgaccacgccc 598

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1451402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba:*
- 2: gb_hgt:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_hgt_hum:*
- 31: em_hgt_inv:*
- 32: em_hgt_other:*
- 33: em_hgt_mus:*
- 34: em_hgt_pln:*
- 35: em_hgt_rod:*
- 36: em_hgt_mam:*
- 37: em_hgt_vrt:*
- 38: em_sy:*
- 39: em_hgt_hum:*
- 40: em_hgt_mus:*
- 41: em_hgt_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81.4	13.6	1372	8	CEL1410
2	79.7	13.2	587	3	TETHEMOGP
3	77.6	13.0	1197	8	CEL1637P
4	76.6	12.8	481	1	AF475938
5	70.4	11.8	146174	1	D90910
6	53.8	9.0	494	3	TETHEMOGT
7	50.8	8.5	3605	1	MKNATNS
8	50.8	8.5	12792	1	AE010389
9	48.2	8.1	1898	10	MUSDELTA
10	48.2	8.1	2330	10	MUSUCRBP
11	48.2	8.1	3041	10	MUSSTRANS01
12	48	8.0	234817	10	AL663048
13	47.6	8.0	110000	2	LMFLCHR32_06
14	47.6	8.0	110000	2	LMFLCHR36_07
15	47.4	7.9	137936	2	CNS08CAX
16	47.4	7.9	144778	2	CNS07EFR
17	47.2	7.9	3014	3	AY084205
18	47	7.9	10486	1	BJDNAKJ
19	46.4	7.8	411	1	MSM49386
20	46.4	7.8	1737	1	AF130980
21	46.4	7.8	16563	1	AE007025
22	46.4	7.8	35377	1	MTCY48
23	46.4	7.8	117080	2	AC098840
24	45.8	7.7	135216	8	AC069145
25	45.8	7.7	144724	8	AC078948
26	45.2	7.6	149117	2	AC120527
27	45	7.5	11401	1	D50308
28	44.8	7.5	5520	3	AC084329
29	44.6	7.5	11034	1	AE005065
30	44.4	7.4	154494	2	AP005070
31	44.2	7.4	11509	1	AE012381
32	44	7.4	12172	1	AE012305
33	44	7.4	79390	2	AC020442
34	44	7.4	84653	3	AC004290
35	44	7.4	167350	2	AC129717
36	44	7.4	180814	3	AC099023
37	44	7.4	256764	3	AE003792
38	43.8	7.3	32274	1	SCC8A
39	43.4	7.3	867	1	AF145230
40	43.4	7.3	110000	2	LMFLCHR36_03
41	43.4	7.3	177466	1	AE01826
42	43.2	7.2	13461	14	AF084543
43	43.2	7.2	195859	14	AF281817
44	43	7.2	201271	2	AC124194
45	43	7.2	213050	1	AL646067

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
CEL1410	C.eugametos	Li410 mRNA.	X72915	GI:437980	haemoglobin; Li410 gene; light-induced expression; nuclear gene.	Chlamydomonas eugametos.	1 (bases 1 to 1372)	Couture, M., Chamberland, H., St-Pierre, B., Lafontaine, J. and Guertin, M.	Nuclear genes encoding chloroplast hemoglobins in the unicellular

